

RESULT 9
TCNA_TRYCR STANDARD: PRT: 1162 AA.
AC P23253;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 20-NOV-2001 (Rel. 40, Last annotation update)
DE SIALIDASE (EC 3.2.1.18) (NEURAMINIDASE) (NA) (MAJOR SURFACE ANTIGEN).
GN TCNA.
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RX SEQUENCE FROM N.A.
RC STRAIN-SILVIO X-10/4;
RA MEDLINE=91277609; PubMed=1711561;
RA Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,
RA Prioli R.P.;
RT "The Trypanosoma cruzi neuraminidase contains sequences similar to
RT bacterial neuraminidases, WTD repeats of the low density lipoprotein
RT receptor, and type III modules of fibronectin."
RL J. Exp. Med. 174:179-191(1991).
RN (2)
RP SUBCELLULAR LOCATION.
RX MEDLINE=91376547; PubMed=1896773;
RA Prioli R.P., Mejia J.S., Ajl T., Alkawa M., Pereira M.E.A.;
RT "Trypanosoma cruzi: localization of neuraminidase on the surface of
RT trypanastigotes."
RL Trop. Med. Parasitol. 42:146-150(1991).
CC -1- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN
CC PARASITE INVASION OF CELLS.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC
CC LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACETYLNEURAMINYL
CC RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACETATED
CC NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,
CC GLYCOLIPIDS OR COLOMINIC ACID.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POSSIBLE).
CC -1- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPAOSTIGOTES, MINIMUM
CC IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.
CC -1- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT
CC DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TCNA
CC PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.
CC -1- SIMILARITY: CONTAINS 3 BNR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M61732; AAA30255.1; -
DR PIR: JH0557; JH0557.
DR HSSP: P29768; IDIL.
DR Interpro: IPR002860; BNR.
DR Pfam: PF02012; BNR. 2.
KW Hydrolase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;
KW Phosphorylation.
FT DOMAIN 1 457 CVS-RICH.
FT REPEAT 23 34 BNR 1.
FT REPEAT 163 174 BNR 2.
FT REPEAT 209 220 BNR 3.
FT DOMAIN 458 588 FIBRONECTIN TYPE-III.
FT DOMAIN 589 1120 44 X 12 AA TANDEM REPEATS, LTR DOMAIN.
FT CARBOHYD 342 342 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1162 AA; 120032 MW; 07049221897C6A40 CRC64;

Query Match 7.3%; Score 530.5; DB 1; Length 1162;
Best Local Similarity 29.4%; Pred. No. 3.7e-14;
Matches 177; Conservative 100; Mismatches 212; Indels 113; Gaps 34;

250

QY 313 TPTTPEKE-PASTTPKEPTPTTTSAPTTTKEKPAPTTTKSAPTTTPKEKPAPTTTP 371
DB 599 TPTSPADSSAHSTPTSPVDS--SAHSTPTSPADSSAHGTPSPVDSAHGTP--PTP 651
QY 372 KE-PAPTTTKEPAPTTTTSAPTTTKE-PAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 418
DB 652 ADSAHGTPSTPVDSSAHSTPTSPVDSAHSTPTSPVDSAHGTPSTPVDSSAHGTPSTP 711
QY 419 KEPTTTP-TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 477
DB 712 VDSAHGTPSTPVDSSAHSTPTSPVDS--AHSTPTSPADSSAHSTPTSPVDS 764
QY 478 PTTTKEPAPTTTTSAPTTTKEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 533
DB 765 HGTPSTPVDSSAHSTPTSP--PADSSAHGTPSTPVDSSAHST--PTSPVDSAHGTPS 817
QY 534 TP-KKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 588
DB 818 TPVDSSAHSTPTSPVDSAHGTPSTPVDSSAHSTPTSPVDSAHST--PTSPVDSAHGTP 876
QY 589 TTP-EKLAPTTTPKEPAPTTTPEELAPTTPEE---PTTTPPEEPA---PTTP-KAAPNT 638
DB 589 TTP-EKLAPTTTPKEPAPTTTPEELAPTTPEE---PTTTPPEEPA---PTTP-KAAPNT 638
APTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 698
ADSSAHST-PTSPADSS--AHSTPTSPVDSAHST 983
3TAPTTTKE-PAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 757
---PTSPADSSAHSTPTSPVDSAHSTPTSPVDSAHSTPTSPVDSAHSTPTSPVDSAHSTPTSP 1033
ISDKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 813
PTSPADSSAHSTPTSPVDSAHSTPTSPVDSAHSTPTSPVDSAHSTPTSPVDSAHSTPTSPVDSAHST 1084
IHSPDESPTP-ELSAEPTPALDNSPREPEVPTTTPPAATKPE 872
DB 1085 DSSAH-STPTSPVDS--AHGTP--STPADSSAHSTPTSPVDSAHSTPTSPVDSAHST 1125
QY 873 MT 874
DB 1126 GT 1127
RESULT 10
ID CPN_DROME STANDARD: PRT: 865 AA.
AC 002910;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE CALPHOTIN.
GN CPN OR CAP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
GN Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CANTON-S.
RX MEDLINE=93165729; PubMed=8094559;
RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT "alphoto: a Drosophila photoreceptor cell calcium-binding protein."
RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
RN [2]
SQ SEQUENCE FROM N.A.
RP STRAIN-CANTON-S.

RX MEDLINE=93165730; PubMed=8434015;
 RA Ballinger D.G., Xue N., Harshman K.D.;
 RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
 RT calcium and contains a leucine zipper."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
 CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
 CC OF CA2 PER MOL. OF PROTEIN.
 CC -1- SUBUNIT: HOMODIMER (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
 CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
 CC COMPOUND EYES AND OCCELLI.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
 CC DEVELOPMENT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L02111; AAA28405.1; -;
 DR EMBL; L05080; AAA28420.1; -;
 DR PIR; A47282; A47282.
 DR FlyBase; FBgn0010218; Cpn.
 KW Calcium-binding.
 FT CONFLICT 36 36 A -> AVAPAVVA (IN REF. 2).
 FT CONFLICT 43 43 I -> T (IN REF. 2).
 FT CONFLICT 64 64 I -> V (IN REF. 2).
 FT CONFLICT 76 76 T -> A (IN REF. 2).
 FT CONFLICT 100 100 P -> PP (IN REF. 2).
 FT CONFLICT 126 127 VO -> AP (IN REF. 2).
 FT CONFLICT 154 154 I -> V (IN REF. 2).
 FT CONFLICT 160 160 S -> T (IN REF. 2).
 FT CONFLICT 534 534 A -> E (IN REF. 2).
 FT CONFLICT 699 699 I -> T (IN REF. 2).
 FT CONFLICT 703 703 V -> L (IN REF. 2).
 FT CONFLICT 721 721 D -> E (IN REF. 2).
 SQ SEQUENCE 865 AA; 84781 MM; 2110417E0B0E7CFE CRC64;
 Query Match 7.3%; Score 530; DB 1; Length 865;
 Best Local Similarity 26.2%; Pred. No. 3,1e-14;
 Matches 233; Conservative 73; Mismatches 433; Indels 152; Gaps 41;

Db 357 AAESIPAPVAVATTVPAT---LAVIDPVTASAVEPELPVLAAPSPVSAVAETPVLAP 413
 Qy 646 TPKEPAPPTPKKPAPPTPKKPAPPTP-KGAPPTLKE-----PAPPTPKKPAPKE 694
 Db 414 V----LPVAAEPPVAAVAAEETPETPAPASAPVIAALDIEVAVIAAPADAPAEAPSA 469
 Qy 695 LAPPTTKPTSTSDKPAPTPKGPAPPTP-----KEPAPTTKEP--ATTTKGTAPT 746
 Db 470 AAPIVSTPP--TTSVPEPTTAPPAVPTPEPIDVSLVSALETVPAPVEVTEVAAVAVD 527
 Qy 747 TLKEPA-----PTPKKPAKEIAPPTT---KGPSTTSDDKAPPT---PKETAPPT 792
 Db 528 APPEAADLIIEPPEPAPPIDLEQTSVPAVEAESTSS--PIPEISLPPEMNAEVS 585
 Qy 793 KEKAPPTPKKPAPPTPEPTTSEVSTPTTKEPTTHKSPD-----ESPPEL 842
 Db 586 EVAAPAPITAPPIP---EPEPSLATPEPIPVAPVIOEAVDAVEVPVETSTSIPT 641
 Qy 843 SAEPPTKALENSPKPEGVPPTTKTPAATKREMTTAKOKTTRDILRTPP--ETTAAPKMT 900
 Db 642 TVE-FPEVAEKVLDPAI--TEAPVTOEPDVANINGAPATEI--TPPAEIVTAAAEVS 697
 Qy 901 KETATTEKTESKITATTTQVST---TTODTTPFKITTLKTTTLA-PKYTTTKKT 953
 Db 698 DIAIPVIDPVPQEIAAVEIETDTKAEVIEGSTIPIEAPVPEVSKYAPVISEAPAA 757
 Qy 954 ITTEINAKKPEETAKPKDRATNSKATTPKPKAPKPKPTSTKPKTMRVYRKPKTPT 1013
 Db 758 EVPTAGNDPNTSVGISSEVPTIAEKREVEVPTS--EIPQSSSPS--DSVPAKTIPL 813
 Qy 1014 PKRMTSTPELNPISRIAEAM---LQTTTRNPQNSKLVEVN-----PKS 1056
 Db 814 LRDLOTTDVSLAATLAIIDAIIGELKDKOKARNOQVMRLCEITKILPPPS 864
 RESULT 11
 FPL_MYTGO
 ID FPL_MYTGO STANDARD; PRT; 872 AA.
 AC 025434;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR (FOOT PROTEIN 1) (MCPPL).
 GN FPL.
 OS Mytilus coruscus (Sea mussel).
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Periomorpha; Mytilloida;
 OC Mytilloida; Mytilidae; Mytilus.
 OC NCBI_TaxID=42192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Foot.
 RX MEDLINE=96394686; PubMed=8798340;
 RA Inoue K., Takeuchi Y., Takeyama S., Yamahata E., Yamazaki F., Odo S.,
 RA Harayama S.;
 RT "Adhesive protein cDNA sequence of the mussel Mytilus coruscus and
 RT its evolutionary implications."
 RT J. Mol. Evol. 43:348-356(1996).
 CC -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
 CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
 CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
 CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
 CC -1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
 CC -1- PTM: PROLINES IN THE REPEAT ARE HYDROXYLATED (SINGLE OR DI-) AND
 CC ALSO TYROSINE (THUS PRODUCING DOPA = 3,4-DIHYDROXYBENILANINE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@slb-sib.ch).

CC -----
 DR EMBL: D63777; BAA09850.1; -
 DR InterPro: IPR002964; Adhesive_plaq.
 DR InterPro: IPR002965; P-rich_extensn.
 DR PRINTS: PRO1216; ADHESIVE.
 DR PRINTS: PRO1217; PRICHTEXTENS.
 KW SIGNAL; Repeat; Hydroxylation.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 872 ADHESIVE PLAQUE MATRIX PROTEIN.
 FT DOMAIN 21 41 NONREPEATIVE LINKER.
 FT DOMAIN 124 872 TANDEM REPEATS OF Y-K-[PS]-K-[IP]-[ST]-Y-
 FT DOMAIN 184 192 P-[PS]-[ST].
 FT DOMAIN 213 221 NONAPEPTIDE 1.
 FT DOMAIN 213 221 NONAPEPTIDE 2.
 SQ SEQUENCE 872 AA; 101677 MW; 98CC70D7C75F3C4 CRC64;

Query Match 7.18; Score 518.5; DB 1; Length 872;
 Best Local Similarity 29.24; Pred. No. 8.5e-14;
 Matches 276; Conservative 80; Mismatches 407; Indels 181; Gaps 54;

OY 180 VYDAGSGLDGDFKVTTPDST--TQH-----NKVSTPKTTAKPIPRESLPPNS 230
 DB 29 VYGSAYSGASAGAYK-TLPQSHPYGSKAHVYKPMNKIIPYI--SKSYPAVYKPKGY 84
 OY 231 DTSKETSLVANKETVETKETTNNKOTSDGKEKTSAKETOSIEKTSADKLAPTSVL 290
 DB 85 YPTNRVQPTGSKTNPPYIKPIAKLSYKAITTYPAYAKSISPPSYK-----HKIT 139
 OY 291 AKPTPKAETTKGPAITTPKEPT--PTTKPEASTTPKEEPTTIKSAPTTPKEAPTTK 349
 DB 140 YPPYK-----PKITYP--PTYKOKPSYPSYKPKTYTPYK-----PKITYPPYK 185
 OY 350 SAPT--TKEPAPT--TKEPAPTTPKEAPTTKEAPTTKSAPTTPKEAPTT--PKK 404
 DB 186 RKSEYTPYKPAITPPYKPKITYP---PYKKRPST---PYKPKTYPPYKPKI 236
 OY 405 PAPTTPKEAPTT--PKEPTPTT--PKEAPTTKEAPTTTPKEAPTT--APKPAAPTTP 457
 DB 237 SYPSIYKPKASYSYSSKSKTTPYTPKISYPTTPKPSYPTTPKPSYPTTPYTP 292
 OY 458 KEAPTTTPKEAPTTTPKEAPTTTPKEAPTTKSAPT--TKEPAPTTPKSAPTTPK 512
 DB 293 K-PKPSYPPYKPKITYPPT--YKPKPSYPPYKOKPSYPPYKSSYPSYKSKTYPP 350
 OY 513 EPSPTTPKEA--PTTPKEAPTTPKP--APT--TPK-----EPAPTTPKPAATTTKKP 562
 DB 351 TYKPKITYPPYKPKPSYPSYKPKKTYSPYKPKITYPPYTPYKPKPSYPSYKKTTPP 410
 OY 563 APAPKEAPTTTPKEAPTT--TTPKKLTTPTEKLAPTTPKEAPTTPELAPTTPEE 617
 DB 411 T-YKPKISYPTTPYKPKASYSYSSKSKTTPYTPKISYPTTPYKPSYPTTPYKPKITYP 469
 OY 618 PT--PTTEEPAPTTPKAAAPTTPK--EPAPTTPKEAPTTTP--KEAPTT--PKETAP 668
 DB 470 PTYKPKPSYPTTPYKPKITYPPYTPYKPKPSYPPYKOKPSYPPYKSSYPSYKSKTYPP 529
 OY 669 TT--PKGTAPTTPKEAPTTTPKPAKELAPTTTKEPTSTSDKAPPT--PKGAPAPTTPK 725
 DB 530 PTYKPKITYPPYK-----PKPSYPSYKPKTYPPYTPYKPKITYPPYTPYKPKASYPYTPYK 583
 OY 726 EPAPTTPKEAPTTTPKGTAPTTPKEAPTTTP--KKPAKELAPTTTKEPTSTSDKPA 781
 DB 584 -----PKITYPPYK--PKPSYTPYKOKPSYPPYKPKSSYPTAPKAKSKTY 628
 OY 782 PTT--PKETAPTTPKEAPTTTPK--KPAPTTPET-----PPTTSEVSTPTTKEPTT 830
 DB 629 PPTPKPTTPYTPYK--PKPSYPSYKPKITYPPYTPYKPKKSSYPPYKSSYPSYQPKKT 687
 OY 831 IHKSPDSTELASPTTPKALENSPKEGVTTPATATKPTPEMTTTPKDKD-----TENDL 886
 DB 688 Y-----PSPYKPKKTYPPYK-----PKISYPTTPK-----TKPSYPSYKPKKTSYPTTPYKPK 735

OY 887 RTPEPTTAAPKMTKEVATTEKTESKITATTTQVSTTTQDTPFKITTLKTTLAPK 946
 DB 736 ISYSTYKAKP-----STPPYKPKPSYSSYKPKIKIRPPYKPKPSYSSYK-----PK 785
 OY 947 VTTTKKTTTTTTEINKNKEETAKPKDRATNSKATTPKPO--KPTKAPKPP--TSTKPKP-T 1001
 DB 786 I-----RYPTTPYKPKP-----SYASSYKPKIRPPYKPKPSYSSYKPKIT 827
 OY 1002 MPYKPKPTT--PPPKKTSMPPELNPSTRALAEMLQTTTPPNQ 1043
 DB 828 YPTTPYKPKISYPTTPYKPKITYPPYKPKISYPPYKPKPSYSSYKPKISYPSQ 871

RESULT 12
 NFH_RAT
 ID NFH_RAT STANDARD; PRT: 831 AA.
 AC P16884; 063368;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).
 GN NFH OR NFH.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=89065087; PubMed=1143606;
 RA Breen K.C., Robinson P.A., Wion D., Anderson B.H.;
 RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
 RT Identification of putative phosphorylation sites.";
 RT FEBS Lett. 241:213-218(1988).
 RN [2]
 RP SEQUENCE OF 37-831 FROM N.A.
 RX MEDLINE=88309090; PubMed=2457365;
 RA Dautigny A., Pham-Dinh D., Rousset C., Felix J.M., Nussbaum J.L.,
 RA Jolles P.;
 RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and
 RT in situ detection.";
 RT Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
 RN [3]
 RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.
 RX MEDLINE=87080760; PubMed=2878828;
 RA Robinson P.A., Wion D., Anderson B.H.;
 RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide
 RT (NF-H).";
 RT FEBS Lett. 209:203-205(1986).
 RN [4]
 RP SEQUENCE OF 318-831 FROM N.A.
 RX MEDLINE=89184647; PubMed=2928342;
 RA Lieberburg I., Splinter N., Snyder S., Anderson J., Goldgaber D.,
 RA Smolowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;
 RT "Cloning of a cDNA encoding the rat high molecular weight
 RT neurofilament peptide (NF-H): developmental and tissue expression in
 RT the rat, and mapping of its human homologue to chromosomes 1 and
 RT 22.";
 RT Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
 RL
 CC -I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -I- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -I- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND

DR	HSSP: P04002; 1MPA.
DR	InterPro: IPR000519; P_trefoll.
DR	Pfam: PF00088; trefoll; 6.
DR	SMART: SM00018; P; 6.
DR	PROSITE: PS00025; P_TREFOLL; 6.
KW	Repeat; Amphibian skin; Glycoprotein; Alternative splicing.
FT	NOL TFR 1 1
FT	DOMAIN 81 144 8 X 8 AA APPROXIMATE TANDEM REPEATS,
FT	
FT	REPEAT 81 88 ALA/THR-RICH.
FT	REPEAT 89 96 1-1.
FT	REPEAT 97 104 1-2.
FT	REPEAT 105 112 1-3.
FT	REPEAT 113 120 1-4.
FT	REPEAT 121 128 1-5.
FT	REPEAT 129 136 1-6.
FT	REPEAT 137 144 1-7.
FT	REPEAT 161 201 1-8.
FT	DOMAIN 218 302 P-TYPE 1.
FT	REPEAT 218 224 8 X APPROXIMATE TANDEM REPEATS, THR-RICH
FT	REPEAT 218 224 2-1.
FT	REPEAT 225 239 2-2.
FT	REPEAT 240 249 2-3.
FT	REPEAT 250 259 2-4.
FT	REPEAT 260 275 2-5.
FT	REPEAT 276 287 2-6.
FT	REPEAT 288 294 2-7.
FT	REPEAT 295 301 2-8.
FT	DOMAIN 306 347 P-TYPE 2.
FT	DOMAIN 353 394 P-TYPE 3.
FT	DOMAIN 402 522 12 X APPROXIMATE TANDEM REPEATS,
FT	
FT	REPEAT 402 411 THR-RICH.
FT	REPEAT 412 419 3-1.
FT	REPEAT 420 431 3-2.
FT	REPEAT 432 443 3-3.
FT	REPEAT 444 453 3-4.
FT	REPEAT 454 460 3-5.
FT	REPEAT 461 472 3-6.
FT	REPEAT 473 479 3-7.
FT	REPEAT 480 491 3-8.
FT	REPEAT 492 498 3-9.
FT	REPEAT 499 515 3-10.
FT	REPEAT 516 522 3-11.
FT	REPEAT 525 566 3-12.
FT	DOMAIN 572 613 P-TYPE 4.
FT	DOMAIN 620 661 P-TYPE 5.
FT	DOMAIN 162 188 P-TYPE 6.
FT	DISULFID 172 187 BY SIMILARITY.
FT	DISULFID 182 199 BY SIMILARITY.
FT	DISULFID 307 333 BY SIMILARITY.
FT	DISULFID 317 332 BY SIMILARITY.
FT	DISULFID 327 344 BY SIMILARITY.
FT	DISULFID 354 380 BY SIMILARITY.
FT	DISULFID 364 379 BY SIMILARITY.
FT	DISULFID 374 391 BY SIMILARITY.
FT	DISULFID 526 552 BY SIMILARITY.
FT	DISULFID 536 551 BY SIMILARITY.
FT	DISULFID 546 563 BY SIMILARITY.
FT	DISULFID 573 599 BY SIMILARITY.
FT	DISULFID 583 598 BY SIMILARITY.
FT	DISULFID 593 610 BY SIMILARITY.
FT	DISULFID 621 647 BY SIMILARITY.
FT	DISULFID 631 646 BY SIMILARITY.
FT	DISULFID 641 658 BY SIMILARITY.
FT	VARIANT 276 276 K -> E.
FT	VARIANT 354 354 C -> R.
FT	VARIANT 415 415 F -> A.
SQ	SEQUENCE 662 AA; 67774 MM; 1085277FE1ED2FD40 CRC64;

Query Match	7.0%;	Score 509.5;	DB 1;	Length 662;
Best Local Similarity	28.1%;	Pred. No. 1.5e-13;		
Matches 223;	Conservative 48;	Mismatches 276;	Indels 247;	Gaps 30;

QY	253	TTNNQSTDSDEKEKETTSAKETQSTSEKTSANDLAPTSYVLAKPRPKAEFTTTKRGALITTKEPK	312
Db	3	TTAAAVATATKDTTAAAGSSAAAEKTTAA-----AGEVSAAPT--AAVAASTEDATY-----	51
QY	313	TPPTPKEPASSTPKKEPTPTTKKSAPTPKKEPATPTTKKSAPTPKKEPATPTTKKEPATPTTK	372
Db	52	-----AAATAAAEYTAAGAPAPTTTAPRTTAAGAPRTTAAATATPTTAAAGAPTTAT	103
QY	373	EPAPPTTKKEPATPTTKKS-APTPKKEPATPTTKKRAPATPKKEPATPTTKKEPTPTTKKEPA-	430
Db	104	GKAPATAAAEVPTTAASKAPRTTAAATHTSAAAPTTAAASAASKERKSTSSSEEHCH	163
QY	431	---PTTKEPAPT---TKREPATPKKRAPATTKK-----EPAPTTKEPATPTTKKE	475
Db	164	VKPSKRMEGSGSKGTTKQ-----CKNNCFDCEKGHOGHICFHARKKRGHSHHEHTTTT	218
QY	476	PSPTPKEPAPTTKKSAPTPTTKKEPATPTTKKSAPTPKKEPSPTTKKEPATPTTKKEPATPTT	535
Db	219	-----APTITQIATITTT---TPTTT-----TTTKKAPPTT-----	245
QY	536	KKRAPTTKKEPATPTTKKEPATPTTKKRAPATPKKEPATPTTKKETAPTKKLLPTTPEKLA	595
Db	246	-----TTTTKAPPTT-----TTTKATTTTT---TPTT-----	270
QY	596	PTTPEKAPPTTPELAPTTPEEPTPTTPEEPATPTTKAAPTPKKEPATPTTKKEPATPTT	655
Db	271	-----TTTTTKATPTTTTTTTTT---TTTTTTTTTTTTTTTT-----	289
QY	656	KEPATPTTKETAPTTKGAPTTKLEKAPATPTTKKPAPELAPTTTKPTSTSDKAPATT	715
Db	290	-----TTTTTKATTTTTTTTTSECKMEPSK-----REDCGYSGITSESCR	328
QY	716	PKG-----TAPTT---PKRAPTTKKEPATPTTKGAPTTKLEKAPATPTTKKPAPELA	765
Db	329	TKGCCFSSISPTQWCHVYTLISOVADCKVPSORVDCGFRGIT----ADQCRKKNCCFSS	384
QY	766	PTTKKGTPTSTSDKAPATPTTKETAPTTKKEPATPTTKKAPATPTTKPTSTSEVPTT	825
Db	385	ISGKWCRCYSISOVA--TTTTPTTTTTPTTTTTTKKATTTT---TTTTTPTTT	436
QY	826	KEPTTIKSPDESTBELASBPTRKALENSPKEBGPVTTTPATPKPEMTTAKDKTERD	885
Db	437	TTTTT-----TTTKA-----TTTTPTTTPTTTTT---KAT---	463
QY	886	LRTPEPTTAAPKTKETATTEKTESKITATTTVOVSTTTQDTPRKITTKITLKTTLAP	945
Db	464	TTTTPTTTTTTTP---TTTTTKAT-----TTTTPTTTTTPTTTTTKATTTTTPTTTTT	510
QY	946	KVTTTKTITTTTLEIMNKPEETAKPKDRATNSKATTPKPKPKAKKPKSTTKPK-----T	1001
Db	511	TTTTTKATTTTTTSECKME-----PSKRADCGYGITSESCRKGCDFSSISPTQWCHVYS	566
QY	1002	MPRVAKKPTPTTPR 1015	
Db	567	LPOVADCKVAPSSR 580	
RESULT 14			
RBP1_HUMAN STANDARD; PRT; 1970 AA.			
AC P24928; 01-MAR-1992 (Rel. 21, Created)			
DT 01-MAR-1992 (Rel. 21, Last sequence update)			
DT 20-AUG-2001 (Rel. 40, Last annotation update)			
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RBP1).			
GN POLR2A.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			

```

RX MEDLINE=92178992; PubMed=1542581;
RA Wintzler M., Acker J., Vicaire S., Vigneron M., Keding C.;
RT "Complete sequence of the human RNA polymerase II largest subunit.";
RL Nucleic Acids Res. 20:910-910(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=95347616; PubMed=7622068;
RA Mita K., Tsuji H., Morimyo M., Takahashi E., Neno M.,
RA Ichimura S., Yamauchi M., Hongo E., Hayashi A.;
RT "The human gene encoding the largest subunit of RNA polymerase II.";
RL Gene 159:285-286(1995).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC THE PHOSPHORYLATION ACTIVATES POL2.
CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND 5.8S GENES.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X63564; CA45125.1; -
DR EMBL: X74874; CA52862.1; -
DR EMBL: X74873; CA52862.1; JOINED.
DR EMBL: X74872; CA52862.1; JOINED.
DR EMBL: X74871; CA52862.1; JOINED.
DR EMBL: X74870; CA52862.1; JOINED.
DR PIR: S21054; S21054.
DR MIM: 180660.
DR InterPro: IPR000684; RNA_polII_repeat.
DR InterPro: IPR000722; RNA_pol_A.
DR InterPro: IPR002879; RNA_pol_A2.
DR Pfam: PF006623; RNA_pol_A: 1.
DR Pfam: PF01854; RNA_pol_A2: 1.
DR PROSITE: PS00115; RNA_POL_II_REPEAT; 43.
KW Transferrase; DNA-directed RNA polymerase; Transcription; zinc; Repeat;
KW DNA-binding; Nuclear protein; Phosphorylation; zinc-finger.
FT ZN_FING 71 87
FT DOMAIN 1590 1958 C2H2-TYPE (POTENTIAL).
FT CONFLICT 1067 1067 W -> L (IN REF. 2).
FT CONFLICT 1449 1449 D -> Y (IN REF. 2).
FT SEQUENCE 1970 AA; 217205 MW; 6876FC25692A657E CRC64;
SQ

```

```

Query Match 6.98; Score 503.5; DB 1; Length 1970;
Best Local Similarity 33.38; Pred. No. 6.4e-13;
Matches 181; Conservative 101; Mismatches 167; Indels 95; Gaps 47;

```

```

QY 312 PPTTPKEPASTTPKE-PPTTIKSAPTTPKEPAPTTTKSAPTTPKE-----PAPTTTKE 365
DB 1507 PSPMGSGSPAMTPWNOGATRAYGAMSPSVSGMTPGAGSPSASASASGSPSPYSAWS 1566
QY 366 PAPTTPKEPAPTTTKPAPTTTSAPTTPKEPAPTTTPKKPAPTTPKP--APTTPKEPTP 423
DB 1567 PTPGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1624
QY 424 TPTTPKEPAPTTTKPAPTTTPKKPAPTTTPKEPAPTTTPKEPAPTTTKPSPPTTPKE 483
DB 1625 TSPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1671

```

```

QY 484 PAPTTPKSAPTTPKEPAPTTTKSAPTTPKEPSPPTTKPAPTTTPKEPAPTTTPKKPAPTTTP 543
DB 1672 -SPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1725
QY 544 KEPAPTTPKEPAPTTTKPAPTTAPKAPAPTTTPKEPAPTTTPKKPAPTTTPKEPAPTTTP 603
DB 1726 SY-SPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1777
QY 604 PPTPELAPPTPEEPPTTPPEPAPTTTPKAAAPNPKPAPTTTPKEPAPTTTPKEPAPTTTPKE 661
DB 1778 PTPSP-NISPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1828
QY 662 TPKEPAPTTTPKGAATPTTPKEPAPTTTPK-KPAKELAPTTTPKEPSTSDKAPAPTTPKGTA 720
DB 1829 SP-SPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1875
QY 721 PPTTPKEPAPTTTPKEPAPTTTPKGAATPTTPKEPAPTTTPKKAPKELAPTTTPKPT-STS 779
DB 1876 PTPSPY-SPTTPKY-SPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1922
QY 780 PAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 837
DB 1923 YSPTSP-SPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1965
QY 838 STPE 841
DB 1966 SDEE 1969

```

```

RESULT 15
NFH_HUMAN STANDARD; PRT; 1020 AA.
ID P12036;
AC 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
GN NFH OR NFH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=88328981; PubMed=3138108;
RA Lees J.F., Smeidman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;
RT "The structure and organization of the human heavy neurofilament
RT subunit (NF-H) and the gene encoding it.";
RL EMBO J. 7:1947-1955(1988).
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC OBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLET K-S-P, NFH IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```
Search completed: April 26, 2002, 16:32:55
Job time: 605 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:25:49 ; Search time 61.21 Seconds
(without alignments)
502.197 Million cell updates/sec

Title: AAF
Perfect score: 7294
Sequence: 1 MAWKPLPYLLLSLVFIQ.....ARATTRSGQLSKWYNCP 1366

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	522	7.2	805	4 US-09-103-429A-4	Sequence 4, App1
2	513	7.0	1837	3 US-08-928-361B-5	Sequence 5, App1
3	506.5	6.9	744	6 5202236-25	Patent No. 5202236
4	499.5	6.8	1721	3 US-08-700-651-5	Sequence 5, App1
5	498.5	6.8	786	4 US-09-103-429A-3	Sequence 3, App1
6	498.5	6.8	1721	3 US-08-928-361B-6	Sequence 6, App1
7	488.5	6.7	826	1 US-07-638-431-2	Sequence 2, App1
8	488.5	6.7	826	5 PCT-US97-00018-2	Sequence 2, App1
9	476.5	6.5	652	6 5202236-13	Patent No. 5202236
10	452	6.2	960	4 US-09-219-849-5	Sequence 5, App1
11	425.5	5.8	1185	4 US-09-041-886-23	Sequence 23, App1
12	424.5	5.8	1867	2 US-08-479-537A-5	Sequence 5, App1
13	424.5	5.8	1867	4 US-09-083-116-5	Sequence 5, App1
14	424.5	5.8	2035	2 US-08-479-537A-2	Sequence 2, App1
15	424.5	5.8	2035	4 US-09-083-116-2	Sequence 2, App1
16	419.5	5.8	2476	2 US-08-276-967-2	Sequence 2, App1
17	417	5.7	829	1 US-08-642-255-132	Sequence 132, App1
18	417	5.7	829	1 US-08-397-633A-53	Sequence 53, App1
19	417	5.7	837	1 US-08-175-155-68	Sequence 68, App1
20	417	5.7	837	1 US-08-477-509B-103	Sequence 103, App1
21	417	5.7	837	1 US-08-642-255-101	Sequence 101, App1
22	417	5.7	837	2 US-08-707-237A-75	Sequence 75, App1
23	417	5.7	837	3 US-08-482-085B-103	Sequence 103, App1
24	417	5.7	837	1 US-08-397-633A-50	Sequence 50, App1
25	413.5	5.7	907	5 PCT-US95-04611A-19	Sequence 19, App1
26	413.5	5.7	907	5 PCT-US95-04611A-19	Sequence 19, App1
27	404.5	5.5	408	1 US-07-609-716-65	Sequence 65, App1

28	404.5	5.5	408	4 US-08-475-411A-65	Sequence 65, App1
29	404.5	5.5	408	4 US-08-478-029A-65	Sequence 65, App1
30	398.5	5.5	682	1 US-08-642-255-126	Sequence 126, App1
31	398.5	5.5	682	1 US-08-397-633A-36	Sequence 36, App1
32	387	5.3	1537	1 US-08-325-267A-2	Sequence 2, App1
33	383.5	5.3	1231	3 US-08-904-263A-4	Sequence 4, App1
34	381	5.2	1848	5 US-08-296-791-6	Sequence 6, App1
35	381	5.2	1848	5 PCT-US95-10661A-6	Sequence 6, App1
36	376	5.2	960	4 US-09-219-849-6	Sequence 6, App1
37	373	5.1	1786	4 US-08-973-462-8	Sequence 8, App1
38	371	5.1	761	2 US-08-707-237A-84	Sequence 84, App1
39	371	5.1	762	1 US-08-642-255-114	Sequence 114, App1
40	371	5.1	762	1 US-08-397-633A-26	Sequence 26, App1
41	370.5	5.1	1064	1 US-08-642-255-62	Sequence 62, App1
42	370.5	5.1	1187	1 US-08-320-559-28	Sequence 28, App1
43	370.5	5.1	1187	5 US-08-545-860D-28	Sequence 28, App1
44	370.5	5.1	1187	5 PCT-US94-04496-28	Sequence 28, App1
45	370.5	5.1	1210	1 US-08-320-559-26	Sequence 26, App1

ALIGNMENTS

```
RESULT 1
US-09-103-429A-4
; Sequence 4, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucln
; TITLE OF INVENTION: CDNA and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
; STREET: 118 No. 6187558th Tlloga
; CITY: Ithaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,429A
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-3628
; TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 805 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; TISSUE TYPE: peritrophic membrane
; US-09-103-429A-4
```

Query Match 7.2%; Score 522; DB 4; Length 805;
Best Local Similarity 27.8%; Pred. No. 5.8e-26;

Matches 226; Conservative 33; Mismatches 211; Indels 342; Gaps 40;

OY 286 TSKYLAAPTKAEATTGKRALTTPEKEPTPTTKKEPASTTPKEPTPTTKSAP-----TTP 340

Db 9 TALGLVARREVSDAEKNPALHEPHDXP--PAEQXXLLPEXEDCKRFYCEXELKFIAP 66

OY 341 KEPAPTT-----TKSAPTTPKKEPAATTTKEPAATTTPEKEPAATTTKEPAATTTK 388

Db 67 KDCAPGIEFKFSQTCYHALAGCTLRGAETTT---QAPATQ--APTTTQ--APTTT 119

OY 389 SAPTTKEPAATTPPKPAATTPKEPAATTPKEPTPTTKKEPAATTTKEPAATTPKEPA 448

Db 120 QAPTTTQATTTT-----QAPTTTQAPTTT-----QATTTQAPTTT----- 156

OY 449 PKKAPTTKEPAATTPKEPAATTTKEPSSTTKKEPAATTTKSAPTTKEPAATTTKSA 508

Db 157 ----QAPTTTQ--APTTTQ--APTTTQ-----APTTTQ--APTTTQ--APTTTQAAA 198

OY 509 TTPKEPSSTTKKEPAATTP--KEPAATTPPKPAATTPKEPAATTPK----- 552

Db 199 TTPPAATTPPAAT--TPAATTPPAATTPPAATTPGVPATTS---APVMPRICELLPNGCAPED 253

OY 553 ----- 560

Db 254 IHLIPHDKXCNLFYQCSNGYFEQRCPREGLYFNRYVQRCDSPANVECDGEISPAVTE 313

OY 561 K-----PA----- 563

Db 314 GNEBEDIDIGLLDNGCPANFEIDMLPBNCRDCKYQCVHGNLVERRCAGTHFSFELQ 373

OY 564 -----PPAP--KEPAAT-----TPKET 578

Db 374 QCDHIEIVGCTLPGESEEVVDVEDACTGWYCEPTIEMLERLNGCPADFSIDLPLPHES 433

OY 579 -----APTPPKKLPTT-----PEKLAATTPKEKPA 604

Db 434 DCGOYLQCVHQHTIARCRPGNLHESPATQSCSESYTAGCVFECDSDSNQSTIAATPAAP 493

OY 605 TTPBELAPTPPEEPTPTTPEEPAPTTPKAAPTKEPAATTPKEPAATTPKEPAATTPK 664

Db 494 TAAPTAAPTAAPTAAPTSTVWPPA--TPPATAPAPVPTTAIPT-----PAPTAAPTAAPTAA 548

OY 665 ETAPTT---PKGATPTLKEPAATTPKKPAKELAPTTKEPTSTSDKRAPTPPGTAP 721

Db 549 PESPTTYVPTTAAPT-----APTT-----APPELPIVTSAPTAAPT--AAPTAAPTAP 598

OY 722 TTPKEPAATTPKEPAATTPKGTAATTLKEPAATTPKKPAKELAPTTTKGPTST--TSDK 780

Db 599 TTAVAPEIPTT-----VTSPTAAPT--AAPAPNT-----TTVAPPTAAPTAP 641

OY 781 APTTPKEPAATTPKEPAATTPKKAPATTPPEMPPTTSEVSTPTTKKEPTTIHKSPEDESAP 840

Db 642 AP-----NTVAPPTAAPT--AAPAPNTTYVPTTAAPTAP-----PIVAH-----AP 685

OY 841 ELASAPTPKALENSPKKEGVPPTTKTPATKPE 872

Db 686 NTAAP-----VTTTSAPATTP--PE 703

RESULT 2

US-08-928-361B-5

Sequence 5, Application US/08928361B

Patent No. 6071518

GENERAL INFORMATION:

APPLICANT: Petersen, Carolyn

TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,

TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS

TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM

TITLE OF INVENTION: SPECIES INFECTIONS

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: PETERS, VERNY, JONES & BIKSA

STREET: 365 Sherman Avenue, Suite 6

CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verny, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480,76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1837 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-5

Query Match	7.0%;	Score 513;	DB 3;	Length 1837;

Best Local Similarity 21.6%; Pred. NO. 5.9e-25;
Matches 307; Conservative 105; Mismatches 539; Indels 470; Gaps 49;

```

0Y 70 HNPSPSSSKARP-----PSGASQIKSTYTKSPD-----PNKKTKKY 110
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 83 HADASTTSAYSAPELDVSGV--PIEPNTRMADVLSIMLEDNSTGVAYDN---TNSI 137
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 111 IESE--EITEEHSVBEONOSSSSSSSSSSSSSTIMIKKSKNSANRELOKILVKNKN 168
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 138 LEGSIAIRSECIYSELNFSTGTFTEDTSMNPV--SITSGELDKPNKQATIGSREC 195
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 169 RKKKPTPKPPYVDAGSLONG--DEKYTTPSTSTQHNKVKSTPKITAKINRPSLP 227
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 196 GKKQ-----GISIDSTGFRDS-----ITGLPTDYPCNP 226
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 228 PMSDTSKETSLVNKET-----TVELKETTTNKOTST---DGK-----EKTTSKE 271
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 227 FNPVGNLYSRSTGKTIPMTYAGVRSNETKTEESANTNELLVPKINAPCNSNSFQO 286
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 272 TQSIKTSKADLAPISKVLAKPTKPAENTTKCPALTTPEPTPTPKBPKEAPTTKPEPPT 331
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 287 VOIEFMGSKVYIPYKCVGKHHTTTTTT-----TTTTTTTTTTTTTTTTTTTTTTT 341
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 332 TIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKKEBAPTTPKKEBAPTTTKSAP 391
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 342 TTTTTTTT-----TTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTT 395
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 392 TTPKPEAPTTPKKPAPTTPKKEBAPTTPKKEPPTTKKEBAPTTKEBAPTTPKKEBAPTAPK 451
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 396 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTT 455
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 452 PAPTTPKPEAPTTKKEBAPTTTKESPTTPKPEAPTTTKSADPTTKKEBAPTTTKSAPTP 511
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 456 PTTTT-----TTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTKKPTTTTTTTTTTTTTT 509
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 512 KKPSPPTTKKEBAPTTPKKEBAPTTPK----- 537
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 510 KKPTTTTTATTTTTTSEESVTKPDMCWELENGCEAKGATYVGVIKDGRIENGMAT 569
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

QY 538 ----- 537
Db 570 MIPNDTHVRFKVDYNTISVRCGAGKLEFPDRSLDFIPVAGHNSCIIYGV 629
QY 538 -----PARTPKE-----PARTPKEPARTTKK 561
Db 630 GGGKIHVSYPGSKDVSLISAPIQCELENEYCDCTAKYGAISHGYQTSADFTVTTAK 689
QY 562 PAPAPKAPAPPTTKEAPPTPKLPTTPEKLAPTTPEKAPPTPEAPPTPEEPTT 621
Db 690 PTTT---TGAPGPTTTTGSFSPKPTTTTKATTT-----TILNPTTTTQKPTT 741
QY 622 TPPEAPPTPKAAPNTPEKAPPTTPEKAPPTTPEKAPPTTPEKAPPTTKE 681
Db 742 T-----T---KVGKPPATTTTTLKPIVTTTTKATTTTTPPT-----TTTTKR 787
QY 682 PAPTPPKKPAK---ELAPTTKEPTS-----TSDKAPPTTKG-TAP 721
Db 788 DEMTTTTPLPDIGDIDITPIPIEKMLDKYRMIVDYNGLLSDNDEPIGSOAGIAD 847
QY 722 TPKEPA-----PTTPEKAPPTPK----- 741
Db 848 TSNEFPVQTHKSTGLPLDPMWGLPDPKSGNLVPTNQTMSGISVSLAKNLTVDIDE 907
QY 742 --GTAPTTLK---EPAPTPPKKPAKELAPTT-----KGPSTTSKDP 780
Db 908 TYGPIIDLTLGYPLDPVSLIPFNPETGELFDPISDEIMNGTIAGIVSISASESLSOKS 967
QY 781 APTTPKE-----TAPTTKEPAPPTPKPAPPTTPTP-----PPTTSVSTPTT 824
Db 968 APIDPATNMVGEFGLNPATGVMIIGSLGPSQTFSPSEIDEGGIIP---EVAANA 1024
QY 825 TKEPTTIHKSPPDESTPE-----LSAPTPKALENSPKKEGVP 861
Db 1025 DKFELSTPSPVESIPREKQKIDISISELMWDIESGRILGOVSKRPISGIAGDLN---P 1060
QY 862 TTKTPAATKEBMTTAKDKTERDLRTPTTTAAPKMTETA---TTEKTESKITA 917
Db 1081 IMKPTPT-----DSVYTKPI---DPTTGLP-FNPPTGHLINPTNNMTDSSFAG 1126
QY 918 TTTQVNTTODTTPFKITTLTKTTTLAPK--VTTTKTITTTTELIMNKPEELAKKDAT- 974
Db 1127 AYKAVSNGIKTDNVYGLPDEITGL-PKDPVSDIPNSTGELVD---PSTGKRIINNYTA 1183
QY 975 -----NSKATTPKQKPTKAPKKEPTSKKPKTMPRVKPKTTPTPKMTS 1019
Db 1184 GIYSKRGRLPIEDENGNLDPSTKLPIDGNOLVNPETNSTVSGTSKPKRPGLPVN 1243
QY 1020 ---TMP-----LNPISRIAEAMLOTTTRNQTNPNSKLVENPKSEDAAG 1061
Db 1244 GGGVAPDEAKDAQDKGKGLIIVPTNSINKDPVNTQVSNITGNI---INP-ETGKV 1297
QY 1062 AEGETPMLLRPHVFMPEVTPD-----MDVLP RVNNGIINP----- 1099
Db 1298 IPGSLPOSILVYPSFNTPOQDEITGKRVDTVYGLPIYPSGELIIDPRTKLPISGVAGDE 1357
QY 1100 MLDSEFNICNG-----PVDGLTTLRNGTLV 1125
Db 1358 ILTEVLNITTDVYGLPIDLETGLPRDPVSGLPQLPGLTLY 1398

```

```

RESULT 3
5202236-25
; Patent No. 5202236
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILIPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456

```

```

; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
; SEQ ID NO: 25
; LENGTH: 744
5202236-25

```

```

Query Match 6.9%; Score 506.5; DB 6; Length 744;
Best Local Similarity 29.9%; Pred. No. 5.2e-25;
Matches 239; Conservative 78; Mismatches 308; Indels 175; Gaps 50;

```

```

QY 212 PKTTAPRINRPSLPRNSDTSKETSITVKNKETTVEKETTITTKKQSTDSKETTSAKE 271
Db 24 PKMTYPPYKPKPSYP-----TYKSKPY-----KPKIT----- 53
QY 272 TOSIEKTSAKDLAPTISKVLKPTPKAETTTKGPALETTTPEKPTTPE-----PASTPKE 327
Db 54 -----YPTTK--AKPS-----YPTTKPKKTYPPYKPKLTPPYKPKP 92
QY 328 PTPPTIKSAPT--TPKEAPPTTKSAPPTPKAPPTTKEAPPTTKEAPPTTKEPA--P 384
Db 93 SYPTTKSKPYKPKKITYPPYKAKPSYPPYKPKKTYPPY--YKPKLTPPYTKPKASYP 151
QY 385 TTTTSAPTTKEPAPPTPKKAPPT--TPKEAPPT---PKEPPTTPEKAPPTKEP--- 436
Db 152 PTVKPKPSYP--PSYKTKKYPPYKPKLTPPYKPKPSYPSYKPKKTYPPYKPKLTP 209
QY 437 -APTPKEPA--TPAPKKA--PTTPEKA--PTTPEKA--PTTKEPS--PTTPEKA 485
Db 210 YPTTKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSY 269
QY 486 -PTTKSAPTTKEPAPPTTKSAPPTKEPSPTTKEPA--PTTPEKA--PTTPEKP 538
Db 270 YPSYTKAKPSY--YPSYTKAKPYTP---PTYKAKPSYPPYKAKPYTKAKPYTKAKP 322
QY 539 A--PTTPEKA--PTTPEKA--PTTTKKA--PTAKKEPAPPTTKEAPPTPKLPTT 590
Db 323 SYPTTKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSY 382
QY 591 PEKLAPTPE--KPAPTPEELAPTTPEEP--PTPEEPA--PTTPEKA--AAPNT---P 639
Db 383 TYKAKPSYPPYKAKPSY--PTYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSY 438
QY 640 KEPAPTTKEBAPT---PEKAPPTTKEAPT---PKGAPTTLKEPAPTTKPAK 693
Db 439 SYPTTKPKISYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 498
QY 694 ELAPTTTKEPTSTSDKAPPTPKGAPPTTPEKAPPTTPEKAPPTTKEGAPTTLKEPA- 752
Db 499 SYPTTKAKPTTPTTYKAKPTTK--ARPSTP---PTYKAKPS-----YPTTKAKPSY 546
QY 753 -PTTKKPAKELAPTTTKGPT--STTSKAPPTTPEKAPPTTPEKAPPTTKEPA--PT 807
Db 547 PPTTKAKPTTK--AKPTTKAKPTVSTYKAKPSYPTTKAKPSY--PTYKAKPSYPT 601
QY 808 ---TPETPPTTSVSTPTTKE---PTTIHSPD-----ESTPELSAEP--PKALENS 854
Db 602 YKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAK 661
QY 855 PKEGVPTTKTPAA-----TKPEMTTAKDKTERDLRTPTTETTTAP-----KMT 900
Db 662 PSYP--PTYKAKPSYPTTKAKPSYPTTKAKPTTKAKPTTKAKPSYPTTKAKPTTKAKPT 719
QY 901 KETATTEKTESKITATTT 920
Db 720 YKAKPTVSTYKAKPTVST 739

```

```

RESULT 4
US-08-700-651-5

```



```

: Sequence 5, Application US/08700651B
: Patent No. 6013882
: GENERAL INFORMATION:
: APPLICANT: PETERSEN, CAROLYN
: APPLICANT: LEECH, JAMES
: APPLICANT: NELSON, RICHARD, C.
: APPLICANT: GUY, JIRI
: TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
: TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
: TITLE OF INVENTION: INFECTIONS
: FILE REFERENCE: 480.19-4(HV)
: CURRENT APPLICATION NUMBER: US/08/700,651B
: EARLIER FILING DATE: 1997-08-14
: EARLIER APPLICATION NUMBER: 08/415,751
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 5
: LENGTH: 1721.
: TYPE: PR1
: ORGANISM: Cryptosporidium parvum
: US-08-700-651-5

```

```

Query Match      6.8%; Score 499.5; DB 3; Length 1721;
Best Local Similarity 21.8%; Pred. No. 4e-24;
Matches 300; Conservative 94; Mismatches 510; Indels 469; Gaps 44;

```

```

QY 93 STTKRSRPPKPKKKKKVIESEITEHVSSENOSSSSSSSSSTIMKIKS-----146
DB 38 SITGSRKDPKPKKAT-----ISGRSCGKQGYSDSGFVDSITGLPT 83
QY 147 --SNKSAANRELQKLKLVKONKRNKTKKPPKPPVDEASGLDNDGFKTTTDTSTQ 204
DB 84 DPNYSCNPFN-----PVTGLVSRSTGKTIPTN-----TYAGVRSN-ETKTEPSANT--129
QY 205 HNKSTSPKITTAKP-----INRPSLPNDSSTKE-----TSLTVN 241
DB 130 YAGYRSNERTKTEPSANTNINLAVPKINAPCENSEFQOQIDMSKVYIPTKCGV 189
QY 242 KETVETKETTNTKOSTDGEKETSAKETOSIEKTSAKDLAPT SKVLAKPTKRAETTT 301
DB 190 KHTTTTTTTTTTTTTTTTTT-----TTTTTTTTTTTTTTTTTTTTTTT 211
QY 302 KGRALTPKKEPTTPKEPASTPKKEPTT IKSAPTTPKEPAPTTPKESAPTTPKEPAPT 361
DB 212 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 271
QY 362 TTKEPAPTTPKEPAPTTPKESAPTTPKESAPTTPKEPAPTTPKESAPTTPKEPAPTTPKEP 421
DB 272 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 331
QY 422 TPTTPKEPAPTTPKEPAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTP 481
DB 332 TTTTTPKKTPTTTTTTTTTT-----TTTTTTTTTTTTTTTTTTTTTTT 377
QY 482 KEPAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTP 513
DB 378 TKPTTTTTTTTTTTTTTKKPTTTTATTTTTTSETESVIKPDCEMCKENKDEBAGATYVG 437
QY 514 -----PSPTT-----TKE 521
DB 438 VIGKDGRIENGMAFTMINDTTHVRFKVDGNTTISVRCRKGAGKLEFPDRSLDFTIP 497
QY 522 P-----APTTPKEPAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTP 545
DB 498 PVAGHNSCIIVGVSGDGKIHVSPPGSKDVSLIS---APIQPSLENEVYCDTCTAYG 553
QY 546 ---DAPTTPKEPAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTP 602
DB 554 AIHSGYQISADPVTITTKKPTTTT--TGAPGQPTTTTGGSSKPTTTTITATTTTTT-----607
QY 603 APTTPEELAPTTPKEPAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTP 662

```

```

DB 608 --TLNPIIITTTOKPPTTT-----TT-----KVAGKRPATTTTTLKPIYTTTTTKATTT 656
QY 663 PKETAPTTPKGTAPTTLKEPAPTTPKPKAPK-----ELAPTTPKAPT-----705
DB 657 TTTTVPPTT-----TTTTKREMTTTTTPLDIDGIEITPPIEMKMDKTRMIDVNSGL 711
QY 706 --TTSDKAPPTTPKG--TAPTTPKEPAPTTPKEPAPTTPKESAPTTPKESAPTTPKESAPTTP 759
DB 712 LIDSNDPEIPESQAGQIADISNLFVOTHKSTGCLPDPMYGLPDRKSGNLVHYTQTM 771
QY 760 APKELAPTTPKGTPTSTSDKAPAPTTPKEPAPTTPKEPAPTTPKESAPTTPKESAPTTPKESAPTTP 819
DB 772 SGLSVSLAKNLTVDIDETYG--LPIDITLGYLDPSLIPFN--BETGELFDPISDEI 827
QY 820 STPT-----TKETPTTHKS-----PDST- 839
DB 828 MNGTIAGIVSGISASESLISOKSALIDPATNMVVEEGGLNPATGVIMPGELPSSDOTQ 887
QY 840 -----PLSAPPTPKALENSPKRPGVPTTKTPATKPEMTTAKDKTTER---884
DB 888 FSPLEDOGILPPEVYAAANADKFLSLIP--PSVP-----ESLPE-----KDKIDSISE 934
QY 885 -----DLRTPETTTAAPKMTKEATTEKTESKIYATT--TQVSTTTQD 929
DB 935 LMVDESGLIGVSKRPIPGSISAGDLNPIKPTQIDSVYGKPIDETTLGLPFNPBGHL 994
QY 930 TTPPKITTLKTTTLAPVYTTTKTITTEIMNKK--BETAKPKO-----RATNSKATT 980
DB 995 INPNNMTMSSFAGAVKAYAVNSGICITDNYGLDVGELTGLPKDPSGIDIPNSTTGEIYVD 1054
QY 981 PKQKPTKAPKPTSTKPKTMP-----RVNPKPTTPPKRMTS 1019
DB 1055 PSTQKPIINNTAGIVSGRPGLPRIEDENGNLFDSTNLPIDGNQNLVNPETNSVSGSTS 1114
QY 1020 --TME-----LNPTSLAEMLQTTTPNQTTPSKL 1049
DB 1115 GTTPKPGIIPVGGGVVPEDEAKDQADKGDGLIVPPTNSINKDPVNTQYSNTTGN--1172
QY 1050 VEVPKSESDAGAGEMPHMLLRPHVMEVYTPD-----MDYLPVPNOGIIINP--1099
DB 1173 --INP--ETGKVIPIGSLPGSINTPSFNTPOQTDITGKPVTVIGLPIPDSTGELIDPAT 1228
QY 1100 -----MLSDETNIONGR-----PVDGLTTLNGLTV 1125
DB 1229 KLPVGSVAGDELLTREVLTITDEVTLGLPIDLETGLPRDPVSGLPQJLNGTIV 1281

```

```

RESULT 5
US-09-103-429A-3
: Sequence 3, Application US/09103429A
: Patent No. 6187558
: GENERAL INFORMATION:
: APPLICANT: Granados, Robert R
: APPLICANT: Wang, Ping
: TITLE OF INVENTION: A No. 6187558e1 Invertebrate Intestinal Nucin
: TITLE OF INVENTION: CDNA and Related Products and Methods
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
: STREET: 118 No. 6187558th Tiooga
: CITY: Ithaca
: STATE: NY
: COUNTRY: USA
: ZIP: 14850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/103,429A
: FILING DATE: 24-JUN-1998

```

[illegible]


```

Db      393 RRRPDKRRNPKRPNKPNPKRNPNNEPSNNKNPN-----EPSNPKRPN-----NEPSN 443
Oy      496 TRNPADPTTTSAPTTTKREPSPTTTKEBPATTPTEPA-PTTKPR-KAPTTRKEBAPTTRKE 553
Db      444 PKNPNP-----NEPSNPKRPN-----EPUNEPSNPNENPNANSPNE--PSNPNE 492
Oy      554 PADTTTKKRAPTPAKPE-AATTREKETAPTTKLGTPTTPEKKLAPTTRKEKPAPTPPEELAP 612
Db      493 PS-----NPNENPNNEPSNPN-----PSNPKK-----PSNPNE--P 523
Oy      613 TTPDEEPTPTTEBPAPTTTKRAARNTPKPE-AATTKEBAPTTRKEBAPTTRKETIAPTP 671
Db      524 SNPN-----PLNPNP-----SNPNENPSNNEPSNPEE--PSNPK-----PSNPNE----- 564
Oy      672 KGTAPTTLKEBAPTTRPKRPARKELATTTTKEPSTSDKRAPTPRKGTAPTTTRKEBAPT 731
Db      565 -----PSNDEEPPDE--PSNPRE-----SNPEEPINPELNPKESNPEESN 606
Oy      732 PKRBAPTTRKGTPTLTKERBAPTTRKRPARKELATTTTGPISTSDKRAPTTREKAPT 791
Db      607 PEKPI-----NPEESNPKPE-----INPEDNENPLIOEBRIEPRANDSVIRI 649
Oy      792 TKREP-----AATPEKRBPATTPETPRPTTSEVSTPTTKKEPTTIHKSDESTPELSAE 845
Db      650 LPIIPCKGNINPSNLPENSDSVEIEPRRDNENSNNTMKSKKNI---PNEPIPSBDG 706
Oy      846 PTPKALENSPK-----EEGVPTTK 864
Db      707 PYGHEERIPIKPHRSNDYVDNNVKNNKDEBEIPNNE 745

RESULT          9
5202236-13
; Patent No. 5202236
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSSBERG,
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILIPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456
; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
; SEQ ID NO:13
; LENGTH: 652
5202236-13

Query Match           6.5%; Score 476.5; DB 6; Length 652;
Best Local Similarity 29.7%; Pred. No. 3.8e-23;
Matches 221; Conservative 71; Mismatches 284; Indels 169; Gaps 45;

Db      212 PKITAAKDINPPSLAPPNDSKTSETSLVNKETEVTETKTTTTTNKQTSIDGKEKTSAKE 271
Oy      11 PKMTIPPYYKKRPSYP-----TYKSKFTY-----KPKIT----- 40
Oy      272 TOSIEKTSAKDLAPNSKVLAKEPTPRAETTTCGRALTTRKEPRTTPKE---PASTTPE 327
Db      41 -----YPRYUK-AKPS-----YPRYUKKRGTPRYKKRLGLVPYUUKRP 79
Oy      328 PPTTIKSGAPT-TREBAPTTTKSASLPTRKEBAPTTRKEBAPTTRKEBAPTTRKEBAPTTRKEBAPT 384
Db      80 SYPRPYKSKPYYKKRITTPYYKAKESTRPYKKRKYTPRT-YKPKLTPPYUUKKASYP 138
Oy      385 TTYSAPPTTRKEBAPTTRKKRAPT-TREBAPT-----PKEBPTTRKEBAPTTRKEBAPT 439
Db      139 PTYKRPSPY-P-PSYKTKTKTTPYTKPKLTLPPTYKKPKPSVPSKMYKDYTPYKKPT 196
```



```

: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 2626
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 23:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1185 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-041-886-23

```

```

Query Match      5.88; Score 425.5; DB 4; Length 1185;
Best Local Similarity 21.88; Pred. No. 1.5e-19;
Matches 240; Conservative 119; Mismatches 437; Indels 303; Gaps 52;

```

```

QY 88 SQTIKSTTKRSPKPKKKKKKKVIESFEITEHSEVSENQESSSSSSSSSTTWKIKSS 147
D 9 SMSNRSGKKKEAPGR-----ELNRSGRASPGGVSTSSSDG-----KAES 50
QY 148 KNSAANRELAKKLVKDKNKKRKKPKPPVVDGAGSLDNGDFKVTTPDTSTQHNK 207
D 51 RQTA-----KKAIVEASTPK--VNKGGR-----EELISEESE 82
QY 208 VSTSPKITTAKPIPNRPSLPNSDTSKETSLTVNKETTVETKETTTNKOTSDGKEKT 267
D 83 ETNAPKTKTEQELRPOSQSDLDLDGRSL--NDGSSPDRIIDQONRST-----P 133
QY 268 SAKETOSIEKTSADKLAPTSKVLAKPKPKAETTKGPAITPKKEPTTP--KEPATTP 325
D 134 SIYSPGSVENDS-----DSSGSLSGPA--RPNRPPLPPSPQRPDSTP 176
QY 326 KEPTPTTIKSAPTTPKBRAPPTTKSAPTTPKEPATTTKKEPATTPKKEPATTP 385
D 177 ROPASP-----EPHPSVPTGYHAPMER--PTSRMGOAR--PARPRHQLYGGT 224
QY 386 --TTKSAPTTPK-----EPAPTPPKKRAPATTPKKEPATTPKKEPTTPK 427
D 225 GGVLSGPRMGKGGGAASVSGRNGCKQNHPRPTTPISVSSSGASGAPTRK---PTTPV 280
QY 428 EPAPTTKEPAR-----TPKKEPARTA--PKKRAPATTPKKEPATTPKKEPTTPK 480
D 281 GGGNLPRAPRANPRNTPRLPRRLRLPLNMSASB---PGLGAORLPGHL--PSRYA 334
QY 481 PKE-----PAPTTTSAPTTTPK--APTTPKSAPTP--KEPSPTTKKEPATTPK 531
D 335 MCGMGGLPRGP---EKGPILAPSPHSLRPASSAPAPARPRPYSSSSSSSAAASSSSS 391
QY 532 PTTPKRAPATTPKKEPATTPK--PAPTTTKKRAPATPKKEPATTPKKEPATTPK 587
D 392 SSSSASBPBAS--QALPSYHNSFPPTSL-----SVSNOP---PKYTOPSLPSQAVMSQ 440
QY 588 --PTTPK--KLAPTTKEKRAPATTPPEELAPTTPEEPTTP----- 621
D 441 GPRPPRYGRLNLSNHNRRPFRSTGASQASTAPRVSTNNHNNQOOOQOOOQOOHNG 500
QY 622 --TPPEAP-----TPKKAARN-----TPKKEPATTPK--APTTPK 657
D 501 NSGPRPGAPRHLLEGSSHNHARYMSPELSGLRYRPRALRPNHNSVYSQAGPNC 560
QY 658 PAPTPKRAPATTPKGAATTLKRAPATTPKPKKAPKELAP--TTTKERTSTSDKRAPY 714
D 561 PAVSSSSNSSSTSGSYPGS--HPSDQOPQAGAPRPFRPVVTVTSSATLSTVIAVTAS 618
QY 715 TPKGATPTTPKKEPATTPKKEPATTPKGAATTLKRAPATTPKPKKAPKELAPTTGAPTS 774
D 619 SPAGYKTASBPDPYGRAPSPGAYKTATPRPYKKGSP-----PSFRGTGTPGYNG--- 670

```

```

QY 775 TTSKAPATTPKKEPATTPKKEPATTPKKEPATTPKKEPATTPKKEPATTPK 834
D 671 -TSPAGGTGTPKPSPTV--GGPLPPAGPS--GLPSLEPPAPAAASPPLS--ATOIKOE 724
QY 835 PDE--STPELSAEPKPALENS--PKERGVTPTTPKATKPEMTTAKDKTTERDLTTP 890
D 725 PAEEYETPE--SPVPRAPSPSPPKVYDVSHASQSRKFKHLDRGFNSCARBDLTFVP 781
QY 891 ETTAAPKMKETATTTTEKT--TESKITATTTQVSTTTODTTPPKITTLKTTTLAPVYT 949
D 782 --LEGSKIAKRRADLVEKVRREAQORAR----- 807
QY 950 TKKTTTTEIMNKKPEELAKPRORATNSKATTPKPKOKTPKAPKKTSTKKTTPRVKPK 1009
D 808 -----EKKERERERERERERERERERERELERSVKLAQEGRAP-----VECP 848
QY 1010 TTPTRKMTSTMPPLNPTSIAEAMLOTTPRNPOTPKSLVEVNPKSEDAAGAGETPHM 1069
D 849 LGPVPHR-----PFEFGSAVA-----TVPYLGPTDTPALRT--LSEYA----- 885
QY 1070 LRPVHFMPEVTPDMYLP 1088
D 886 --REVHMSPGNRNHPFYVP 902

```

```

RESULT 12
US-08-479-537A-5
; Sequence 5, Application US/08479537A
; Patent No. 5861381
; GENERAL INFORMATION:
; APPLICANT: CHAMBON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREVENT, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22133-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,537A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 90/13101
; FILING DATE: 23-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021

```

```

Query Match Similarity      5.8%; Score 424.5; DB 2; Length 1867;
Best Local Similarity      23.7%; Pred. No.2.9e-19;
Matches 253; Conservative    69; Mismatches 511; Indels 241; Gaps 53;

QY   236 TSLTVNKKETVVTBKETTNTNKQTSDGKEKRTTSAKETOSIEKTSARD-LAPTSKVLAKPT 294
      |||          :|||:|||:|||:|||:|||:|||
DB   16 TLVLT-----VTGSGHASSTPGGEKERKSATQRSSSPSTSEKNNAVSMSTSLSLSS 65

QY   295 P-KAETTKGP--ALTTRP-----TPTRPKPASTPKPEPTTIKSNAP 337
      :|||:|||          |||          |||          |||
DB   66 PGSGSSTTGODVTLIAPATERASGSNAIWGCVTSVPYTRPALGSTTPRANHVT---SNP 122

QY   338 TPKEEAPATT-----KSAPTTKPEP-----ADTTKEEAPATTPKKEAPATT 378
      |||:|||          |||          |||          |||
DB   123 --DNKPAPGSTARXAHGVTSAPDXRPXPGSTARXAHGVTSAPDXRPXPOSTARXAHGVTS 180

QY   379 TE--PARPTT-----KSAPTTKEEAPATTPK-----KRAPTTKEEAPATTPKKEEPTT 425
      |||:|||          |||          |||          |||
DB   181 ADXRPRXPGSTARXAHGVTSAPDXRPXPGSTARXAHGVTSAPDXRPXPOSTARXAHGVTS 240

QY   426 PKE--PARPTKPEP-----AEPTTKEEAPATTARK-----KRAPTTKEEAPATTPKKEEAPATT 472
      :|||:|||          |||          |||          |||
DB   241 AFDXRPRXPGSTARXAHGVTSAPDXRPXPGSTARXAHGVTSAPDXRPXPGSTARXAHGVTS 300

QY   473 TEPESTTTTKEEAPTT--TKSAPTTTKEEAPTT-----TKSAPTTTKEESPITTK----E 521
      :|||:|||          |||          |||          |||
DB   301 ABDXP-RXGSTARXAHGVTSAPDXRPXGSTARXAHGVTSAPDXRPXGSTARXAHGV 359

QY   522 PARPTTKEEAPATTPK-----KRAPTTKEEAPATTPK-----EPATTTTKKRAPATTKEEAPATT 573
      :|||:|||          |||          |||          |||
DB   360 SAPDXRPXGSTARXAHGVTSAPDXRPXGSTARXAHGVTSAPDXRPXGSTARXAHGV 419

QY   574 TPKEEAP-----TPRKLIPTTPEKLAPTTPEKRAPATTPEELAPTTPEEPTPTTPEEAPATT 629

```

```

Dh 420 SAPDXRPGSTAPXAHGVTS-----APDXRPGSTAPXAHGVTSAPDXRPX- ---PGST 471
QY 630 TPBK-----AAPNPKPEAPPTPKPEAPPTPKPEAPPTPKETAPTTPKGT-APTTLKEPAP 684
Dh 472 APXAHGVTSAPDXRPGSTAPXAHGVTSAPDXRP- XPGSTAPXAHGVTSAPDXRPGXS 530
QY 685 TTP-----KKPAPKELAPTTKEPTSTTSOKPAP--TTPKGTAFTTPKEAPPTP 732
Dh 531 TAPXAHGVTSAPDXRPGSTAP- XAHGVTSAPDXRPGSTAPXAHGVTS---APDXR 585
QY 733 KEPAPTPKG-----TAPTTLKEAPPTP-----KKPAPKELAPTTKGPTSTSD 778
Dh 586 PXPSTAPXAHGVTSAPDXRPGSTAPXAHGVTSAPDXRPGSTAP- XAHGVTSAPDX 644
QY 779 KPAP--TTPK-----ETAPTPKEPAPTTP-----KKPAPTTPTPTTSEVSTP 822
Dh 645 RPXGSTAPXAHGVTSAPDXRPGSTAPXAHGVTSAPDXRPGS--TAPXAHGVTSAP 702
QY 823 TTYKEPT-----IHSDESTPELSAPPTPKALENSPKREVYTTTPAA----- 868
Dh 703 DXRXPGSTAPXAHGVTSAPDXRPGSTAPXAHGVTSAPDXRPGSTAPXAHGVTSAP 762
QY 869 -TKPEMTTAKD---TTEEDLRTPTETTPAAKMTKEATTEKTESKIATTTQVTS 924
Dh 763 DXRXPGSTAPXAHGVTSAPDXRPGST--APXAHGVTSAPDXRPGSTAPXAHGVTS 820
QY 925 TTTODTTPFKITLTKTTTLAPKVTYTTKITTITTELMNKPBEATAKPKDRATNSKATTPKQ 984
Dh 821 APDXRPGX-----STAPXAHGV-----SAPDXRPGSTAPXAHGVTSAPDXRPGX- 868
QY 985 KPAPAKPKPTSTKKPKMTPRBRKKTTPTRKMTSTMPBELNPTSIATAMLOTTTRPOT 1044
Dh 869 -GSTAPXAHGVTSAPDXR- XPGSTAPXAHGVTS- PDXRPX-----PGST 911
QY 1045 -PNSKLEVENKSEADAGABEETPHMLLRPHVMEVTPDMDYLPRVNOGIIINPMUSD 1100
Dh 912 APXAHGVTSAPDXRPGSTAPXAH-----GVTSAPDXRPGSTA---PXAHG 957
QY 1104 ETNIONCKPVDGLTT-LRNGILVFRGHYFMWLSPPSPERARILEWGCISPTD 1158
Dh 958 TSAPDXRPGSTAPXAHGVTS-----PDXRPGSTAPXAHGVTSAPD 1003

RESULT 13
US-09-083-116-5
; Sequence 5, Application US/09083116
; Patent No. 6203795
; GENERAL INFORMATION:
; APPLICANT: CHAMON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LAJHE, Richard
; APPLICANT: HAREVENT, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22133-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083.116
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```



```

APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIORITY DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIORITY DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIORITY DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Rodin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 138..1727
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
FEATURE: repeats varies from 1 to 40."
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-5

```

Query Match 5.8%; Score 424.5; DB 4; Length 1867;
 Best Local Similarity 23.7%; Pred. No. 2.9e-19;
 Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

```

236 TSLTVKKEVETKETTNTKOTSTGKEKTSKETSIAKSD-LAPTSKVLAKPT 294
16 TVLTV-----VTSGHASTPGGEKETSATGROSSVSSTKKNVMSVLSHS 65
295 P-KAETTTKGP--ALTTPKEP-----TPTTPKEPSTYKKEPTPTTINSAP 337
66 PGGSSSTTGQGVTLAPLTPERASGAATWGDVTSVPTPALGSTTPRAHDVT---SAP 122
338 TTPKPAAPTTT-----KSAPTTTPKEP-----APTTPKEPAPTTPKEPAPT 378
123 --DNKPAAGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVT 180

```

```

379 TKE--PAPTTT-----KSAPTTTPKEPAPTTPK-----KRAPTTTPKEPAPTTPKEPPTT 425
181 APDXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVT 240
426 PKE--PAPTTKEP-----APTTPKEPAPTTPK-----KRAPTTTPKEPAPTTPKEPAPT 472
241 APDXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVT 300
473 TKEPSTTPKEPAPTTP--TSAPTTTPKEPAPT-----TSAPTTTPKEPAPTTPK-----E 521
301 APDXRP--XPGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVT 359
522 PAPTTTPKEPAPTTPK-----KRAPTTTPKEPAPTTPK-----EPAPTTTPKRAPTTTPKEPAPT 573
360 SAPDXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVT 419
574 TPKETAP-----TTPKLTPTTPKEKLAAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 629
420 SAPDXRXPXGSTAPXAHGVTSA-----APDXRXPXGSTAPXAHGVTSAPODXRXPX-----PGST 471
630 TPKA-----AAPNTPKPEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 684
472 APXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPX 530
685 TTP-----KKPAKELAPTTTKEPTSTSDKPAAP--TTPKGTAPTTTPKEPAPTTP 732
531 TAPXAHGVTSAPODXRXPXGSTAP--XAHGVTSAPODXRXPXGSTAPXAHGVTSA-----APDXR 585
733 KEPAPTTPKG-----TAPTTLKBPAPTTP-----KKPAKELAPTTTKEPTSTSD 778
586 PXPGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAP--XAHGVTSAPODX 644
779 KPAP--TTPK-----ETAPTTTPKEPAPTTP-----KKPAPTTPPTTPPTTSEVSTP 822
645 RXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAP--TAPXAHGVTSA 702
823 TTPKEPT-----IHSPESTPELSAEPPTPKALENSPKPEGVPTTKTPAA----- 868
703 DXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSA 762
869 -TKPEMTTTPAKDK--TTERDLRTTPEPTTAPPKMTKETATTTKTESKTTATTTQVTS 924
763 DXRXPXGSTAPXAHGVTSAPODXRXPXGST--APXAHGVTSAPODXRXPXGSTAPXAHGVT 820
925 TTTQDTPPKITTLTTLAPLAVTTTKTITTTTETIMNKPEETAPKDRATNSKATTPK 984
821 APDXRXPXG-----STAPXAHGVT-----SAPDXRXPXGSTAPXAHGVTSAPODXRXP 868
985 KPTKAPKKTSTKRPKTPRVKPKTTPTPPKMTSTPELNTPTSRILEMLOTTRPNOT 1044
869 -GSTAPXAHGVTSAPODXR--XPGSTAPXAHGVTSA--PODXRXP-----PGST 911
1045 -PNSKLEVNKSSDAGAGETPHEMLLRPHVPEVTTPMDYLPVPNOGIIINPLMSD 1103
912 APXAHGVTSAPODXRXPXGSTAPXAH-----GVTSAPDXRXPXGSTA-----PXA 957
1104 ETNICKGKPVQGLTTP--LRNGTLVAFRGHYFMNLSPEPSPARITIEWGCTSPID 1158
958 VTSAPDXRXPXGSTAPXAHGVTSA-----PDXRXPXGSTAPXAHGVTSAPODX 1003

```

RESULT 14
 US-08-479-537A-2
 Sequence 2, Application US/08479537A
 Patent No. 5861381
 GENERAL INFORMATION:
 APPLICANT: CHAMON, Pierre
 APPLICANT: KIENY, Marie-Paule
 APPLICANT: LATHEY, Richard
 APPLICANT: HAREDEVINT, Maria
 TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
 TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION:
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION:
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION:
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACG, ACGT,
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION:
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCG, CCGT,
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION:
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."

US-08-479-537A-2

Query Match	5.8%;	Score 424.5;	DB 2;	Length 2035;
Best local Similarity	23.7%;	Pred. No. 3.2e-19;		
Matches 255;	Conservative	69;	Mismatches 511;	Indels 241;
				Gaps 53

QY	236	TSUVNKEETIVKRETTTTHKOSTGCKRKTTSAAETOSIEKYSKD-LATSTSVLAKRP	234
Db	16	TVLV-----VSGSHASTGCGKEKFSATQORSSVSPSTKNAVSTSVLSHS	65
QY	295	P-KAETTTKGD--ALITPKP-----TPITPKEPASTPKKEPTTTIKSA	337
Db	66	POGSSSTIGODVTLAPATEPAGSAAATWGODVSVPIRALGSTTPPHADVT---SAP	122
QY	338	TTPKPAPTTT-----KSAPITPKP-----APITTKEPAPITPKEPAPT	378
Db	123	--DNKPAAGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVT	180
QY	379	TKE--PAPTT-----KSAPITPKEPAPITPK-----KRAPITPKEPAPITPKEPIT	425
Db	181	APDXRPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVT	240
QY	426	PKE--PAPITKEP-----APITPKPAPITAK-----KRAPITPKEPAPITPKEPAPT	472
Db	241	APDXRPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVT	300
QY	473	TKESPITPKEPAPIT--TKSAPITTTKEPAPT---TKSAPITPKESPITTK-----E	521
Db	301	APDXRP--XGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVT	359
QY	522	PAPITPKEPAPITPK-----KRAPITPKEPAPITPK-----EPAPITTKKAPAPKREPAPT	573
Db	360	SAPDXRPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVT	419
QY	574	TPKETAP-----TTPKLTPTTPEKLAPITPKEPAPITPEELAPITPEEPITPTTPEEPAPT	629
Db	420	SAPDXRPXGSTAPXAHGVT---APDXRPXGSTAPXAHGVTSAPOXRPX---PGST	471
QY	630	TPKA-----AAPITPKEPAPITTKEPAPITTKEPAPITTKEPAPITTKPGT-APTTLKEPAP	684
Db	472	APXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPX--XPGSTAPXAHGVTSAPOXRPXGS	530
QY	685	TTP-----KKRAPKELAPITTKEDPTSTSDKAP--TTPKGAPTTKEPAPITPK	732
Db	531	TAPXAHGVTSAPOXRPXGSTAP--XAHGVTSAPOXRPXGSTAPXAHGVT---APDXR	585
QY	733	KEPAPITPKG---TAPITTLKEPAPITPK-----KKRAPKELAPITTKEDPTSTSD	778
Db	586	PXPGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXPGSTAP--XAHGVTSAPOX	644
QY	779	KPAP--TTPK-----ETAPITPKEPAPITPK-----KKRAPITPEPPTTSEVSTP	822
Db	645	RFXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGGS--TAPXAHGVTSAPO	702
QY	823	TTTTKEPTT-----IHKSPDESTRLSAPITPKALESPKPEGVPTTKTPAA-----	868
Db	703	DXRPXPGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPO	762
QY	869	--TKPEMTTAKDK-----TTERDLKTPPEPTTAPAKMTKEGATTTTEKTESKITATTOVTS	924
Db	763	DXRPXPGSTAPXAHGVTSAPOXRPXGGS--APXAHGVTSAPOXRPXGSTAPXAHGVT	820
QY	925	TTTTODTTPKLTTLKTTTLAPKYTTTKKTTTTTEILIMNKPEELAKPODATNSKATTPKPO	984
Db	821	APDXRPXG-----STAPXAHGVT-----SAPDXRPXGSTAPXAHGVTSAPOXRPX-	868
QY	985	KPTAPAKKPTSTKKKPTMPVRKPKTPTPKMTSTIMELNPTSKIAELMOTTTRENOT	1044
Db	869	-GSTAPXAHGVTSAPOXRP--XGSTAPXAHGVTSA--POXRPX-----PGST	911
QY	1045	-PNSKLVENVNKSDDAGABGEPTRHMLRHYHVMPEVTPTDMDUITYLPRVNOGIIINPMLSD	1103

Db 912 APXAHGVSAPDXRXPXGSTAPXAH-----GVTSAPDXRXPXGSTA-----PXAHG 957
QY 1104 ETNINCGKRPVDELTT-LRNGTLVAFRGHYFWMLSPSPSPARRITEVWGISPID 1158
Db 958 VTSAPDXRXPXGSTAPXAHGVSIA-----PDXRXPXGSTAPXAHGVSAPD 1003

RESULT 15
US-09-083-116-2
Sequence 2, Application US/09083116
Patent No. 6203795
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierie
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUVENI, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,116
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 01753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "amino acid 134 is X1 - Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro - CCT,

OTHER INFORMATION: CCC, CCA, or CCG; and Ala - GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y - Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr - ACT, ACC,
OTHER INFORMATION: or ACG; and Asn - AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 - Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro - CCT, CCC,
OTHER INFORMATION: or CCG; and Ala - GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-2

Query Match 5.8%; Score 424.5; DB 4; Length 2035;
Best Local Similarity 23.7%; Pred. No. 3.2e-19;
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 236 TSLVNNKETTVEETTTKQSTDKKETSKEKTSKQTSKETSAND-LAPTSLVLAAPT 294
Db 16 TVLTV-----VSGSHASTPGGKERETATORSVSSPEKNAVMTSSVLSHS 65
QY 295 P-KAETTKGP--ALTPKRP-----RTTPKREPASTPKKEPTTTKISAP 337
Db 66 PGSSSTTGODVYLAATEPAGSAAATNGODVTSVVTTPALGSTTPPAHDVT--SAP 122
QY 338 TTKEPAPTTT-----KSAPTTKRP-----APTTPKEPAPTTKEPAPT 378
Db 123 --DNKPAFGSTAPXAHGVSAPDXRXPXGSTAPXAHGVSAPDXRXPXGSTAPXAHGVS 180
QY 379 TKE--PAPTT-----KSAPTTKREPAPTTK-----KAPTTPKEPAPTTKEPPTT 425
Db 181 APDXRXPXGSTAPXAHGVSAPDXRXPXGSTAPXAHGVSAPDXRXPXGSTAPXAHGVS 240
QY 426 PKE--PAPTTKP-----APTTPKEPAPTAPE--KPAPTTPKEPAPTTKEPAPT 472
Db 241 APDXRXPXGSTAPXAHGVSAPDXRXPXGSTAPXAHGVSAPDXRXPXGSTAPXAHGVS 300
QY 473 TKESPPTTKEPAPT--TKSAPTTTKEPAPT--TKSAPTTTKEPAPT--E 521
Db 301 APDXRP--XPGSTAPXAHGVSAPDXRXPXGSTAPXAHGVSAPDXRXPXGSTAPXAHGVS 359
QY 522 PAPTTPKEPAPTTPK-----KPAPTTPKEPAPTTPK-----EPAPTTPKPAAPTAPEAPT 573
Db 360 SAPDXRXPXGSTAPXAHGVSAPDXRXPXGSTAPXAHGVSAPDXRXPXGSTAPXAHGVS 419
QY 574 TPKEAP--TTPKLTPTPEKLAPTTPPEKPAPTTPPELAAPTTPPEEPPTTPPEPAPT 629
Db 420 SAPDXRXPXGSTAPXAHGVS--APDXRXPXGSTAPXAHGVSAPDXRXPXGSTAPXAHGVS 471
QY 630 TPKA-----AAPNTKREPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPT 684
Db 472 APXAHGVSAPDXRXPXGSTAPXAHGVSAPDXRXPXGSTAPXAHGVSAPDXRXPXG 530
QY 685 TTP-----KKPAPELAAPTTPKEPTSTTSOKPAE--TTPKGTAPTTTKEPAPTTP 732
Db 531 TAPXAHGVSAPDXRXPXGSTAP--XAHGVSAPDXRXPXGSTAPXAHGVS--APDXR 585
QY 733 KEPAPTTPG-----TAPTTLKEPAPTTP-----KKPAPELAAPTTPKTGSTSTSD 778
Db 586 PXPGSTAPXAHGVSAPDXRXPXGSTAPXAHGVSAPDXRXPXGSTAP--XAHGVSAPDX 644
QY 779 KPAP--TTPK-----ETAPTTTKEPAPTTP-----KKAPPTTPEPTTPPTTSEVSTP 822
Db 645 RXPXGSTAPXAHGVSAPDXRXPXGSTAPXAHGVSAPDXRXPXG--TAPXAHGVSAP 702

```

QY 823 TTKEPT-----IKSPDESTPELSAEPTEKALENSPEBGPVTTKTPAA----- 868
Db 703 DXRPXPGSTAPXAHGVTSA PDXRPXPGSTAPXAHGVTSA PDXRPXPGSTAPXAHGVTSA PD 762
QY 869 -TRPEMTTAKDK---TTERDLRTTPETTAARPKMTKETATTEKTESKITATTOYTS 924
Db 763 DXRPXPGSTAPXAHGVTSA PDXRPXPGSTAPXAHGVTSA PDXRPXPGSTAPXAHGVTSA 820
QY 925 TTQODTTPFKITTLKTTTLPKVTYTTTKKTTTTEIMNKPEETAKPKDRATNSKATTPKPQ 984
Db 821 APDXRPXPG-----STAPXAHGVT-----SAPDXRPXPGSTAPXAHGVTSA PDXRPX 868
QY 985 KPTKAPKPKPTSTKKPKTMPRVKPKTTPTRKMTSTMPPELNPISRIAEAMLQTTTRPNQT 1044
Db 869 -GSTAPXAHGVTSA PDXRPXPGSTAPXAHGVTSA PDXRPXPGSTAPXAHGVTSA PD 911
QY 1045 -PNSKLYEVNPKSEDAGABGETPHMLLRPHVEMPEVTPDMDYLP RVYVNOGIIINPMLSD 1103
Db 912 APXAHGVTSA PDXRPXPGSTAPXAHGVTSA PDXRPXPGSTAPXAHGVTSA PD 957
QY 1104 ETNIONGKRPVVDGLT-LRNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSPID 1158
Db 958 VTSAPDXRPXPGSTAPXAHGVTSA PDXRPXPGSTAPXAHGVTSA PD 1003

```

Search completed: April 26, 2002, 16:26:07
 Job time: 507 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:32:20 ; Search time 114.61 Seconds
(without alignments)
904.577 Million cell updates/sec

Title: AA7
Perfect score: 7323
Sequence: 1 MAMKTLPIYLILLISVEVIO.....ARATTSRGQTLSKWNVNC 1361

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR.68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1088.5	14.9	3020	2 A43932	mucin 2 precursor,
2	950	13.0	1664	2 T18262	S-layer protein -
3	867	11.8	1489	2 T31108	cyst germination s
4	835	11.4	1274	2 T16251	hypothetical prote
5	828.5	11.3	2187	2 T30826	nascent polypeptid
6	800	10.9	1188	2 S49915	extensin-like prot
7	795.5	10.9	1367	2 S48478	glucan 1,4-alpha-g
8	704.5	9.6	1151	2 T18535	high molecular mas
9	674	9.2	1229	2 T25697	hypothetical prote
10	672	9.2	1344	1 A35175	mucin 1 precursor,
11	670	9.1	7962	2 T38346	elastic titlin - hu
12	668	9.1	3507	2 T34513	hypothetical prote
13	663	9.1	3570	2 T45025	mucin MUC5B, trach
14	646	8.8	990	2 T51622	nucleolar phosphop
15	633	8.6	489	2 T11622	extensin class 1 p
16	632	8.6	761	2 C84672	hypothetical prote
17	632	8.6	971	2 T19431	hypothetical prote
18	626	8.5	6642	2 T29757	protein UNC-69 - C
19	622	8.5	839	2 T75518	hypothetical prote
20	607.5	8.3	801	2 T29018	hypothetical prote
21	607.5	8.3	924	2 S27923	gene Lp3 protein -
22	605	8.3	379	2 S50125	larval glue protei
23	568	7.8	2232	2 T34434	hypophenolic adhes
24	566.5	7.7	875	2 S23760	ascites stialoglyco
25	565	7.7	1630	2 A53577	hydroxyproline-ric
26	559.5	7.6	350	2 S22456	hypothetical prote
27	556.5	7.6	856	2 T15543	hydroxyproline-ric
28	551	7.5	620	2 S06733	calphostin - fruit
29	545	7.4	873	2 A47283	

30	542	7.4	369	2 S20500	hydroxyproline-ric
31	538.5	7.4	416	2 J00465	extensin precursor
32	535	7.3	756	2 T27642	hypothetical prote
33	533	7.3	1087	1 QEMSH	neurofilament trip
34	530.5	7.2	1162	2 JH0557	exo-alpha-stallidas
35	530	7.2	865	2 A47282	calcium-binding pr
36	528	7.2	3534	2 T42567	tegment protein 2
37	527.5	7.2	1459	2 T32271	hypothetical prote
38	522	7.1	328	2 J00985	hydroxyproline-ric
39	519	7.1	813	2 S70795	vsaA protein precu
40	518	7.1	866	2 T45462	membrane glycoprot
41	518	7.1	1072	1 A37221	neurofilament trip
42	514	7.0	1611	2 T38236	hypothetical prote
43	512.5	7.0	867	2 T45463	membrane glycoprot
44	510.5	7.0	1832	2 T31113	mucin-like glycopr
45	509.5	7.0	662	2 A45155	mucin FTM-C.1 - Af

ALIGNMENTS

RESULT 1
A43932
mucin 2 precursor, intestinal - human (fragments)
N:Alternate names: mucin SMUC-41
C:Species: Homo sapiens (man)
C>Date: 10-Mar-1993 #sequence, revision 12-Apr-1996 #text change 05-Nov-1999
C:Accession: A49963; A45106; B45106; A43932; B33532; A61257; P00328; P00329
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t
A:Reference number: A49963; MUID:94132002
A:Accession: A49963
A:Molecule type: mRNA
A:Residues: 1-639 <GU1>
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up
A:Reference number: A45106; MUID:93016075
A:Accession: A45106
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 626-1895 <GU2>
A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396
A:Note: sequence extracted from NCBI Backbone (NCBIP:116706)
A:Accession: B45106
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GU3>
A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398
A:Experimental source: colon
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J. Clin. Invest. 86, 1005-1013, 1991
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
A:Reference number: A43932; MUID:91358717
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
A:Note: sequence inconsistent with the nucleotide translation
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi
A:Reference number: A33532; MUID:89197956
A:Accession: B33532
A:Molecule type: mRNA
A:Residues: 1916-2193 <GU4>
A:Cross-references: GB:M2405; NID:g188873; PIDN:AAA36334.1; PID:g188874
A:Experimental source: intestine
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

J. Clin. Invest 87, 77-82, 1991

A:Title: Human bronchus and intestine express the same mucin gene.

A:Reference number: A61257; MUID:91086481

A:Accession: A61257

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 'r', 1925-1948, 'rTS', 1952-1954 <I>AN>

A:Experimental source: bronchus

R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner Biochem. Biophys. Res. Commun. 183, 821-828, 1992

A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminus

A:Reference number: P00328; MUID:92198477

A:Accession: P00328

A:Molecule type: mRNA

A:Residues: 2328-2468 <XUG>

A:Cross-references: GB:M85623

A:Experimental source: small intestine

A:Accession: P00329

A:Molecule type: protein

A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>

A:Cross-references: GB:M85623

A:Experimental source: small intestine

A:Gene: GDB:MUC2

A:Cross-references: GDB:120203; OMIM:158370

A:Map position: 11p15.5-11p15.5

C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology: von Willebrand factor type C repeat homology <WMC>

F:2766-2834/Domain: von Willebrand factor type C repeat homology <WMC>

Query Match 14.9%; Score 1088.5; DB 2; Length 3020;

Best local similarity 28.2%; Pred. No. 2.7e-38;

Matches 379; Conservative 98; Mismatches 514; Indels 355; Gaps 51;

43 TCNCDY--NQAHMECCPDKKRYCTAELSCKGKCFSEFEGRCDDAQCCKYDKCCPDY 100

1079 SCGCDTGDECECSAVASIAQECTKEGAC-----VFMTTPDL-CPIFCDDYNN---PPH 1128

101 ESFCAEVHNPTSPSSKAPPPSGASQT-----KSTTKSPK--PPNKKTKKVIIESE 153

1129 E--CEMYEPCGNNSFTCRTINGHNSINISVYLEGGYPCPKDRPIYEDLKKCVIADK 1186

154 I-TLVKNNKKNRKKKPKPK-----PPVY--DEAGSGLD---NGDF--KVTTPD 194

1187 CGCYVEPTNHPGASVPEPEIKCSVCYVNSQYVCREBEKKILNQDGAFCYWEIGCPN 1246

195 TSTYQH--NKVSTSPKTIITAKPINRPSLPSPNSDTSKESLIYNKETVETKET--TTT 249

1247 GTVEKHNFGISITRRPSLITFTTITLPTP-----SFPTTTTTPTSTVLSLT 1298

250 NK-----QTSFGKEKITS--AKETOSIEKTSARD----- 277

1299 PKLCLMSDWINEDHPSSGSDGDREPDVCGAPEDIECSVNDPHLSLEGHQKWCOD 1358

278 -----IAPTSKVLAKTPPAETTTKGP 300

1359 VSVGICKENEDQFGNGPGLCYDKIRVNCMPMDKCIITPSPPPTTSPPTTTTLP 1418

301 LTTKEPPTTPKPEASTTPKEPPTTIKSAF--TPKPEAPATTKSATTPKEPAATT 359

1419 TTPSPPTTTPPTTTPPTTSPPTTTTTPLDPTTPSPISITTTTTPPTTTPSP--PTTP 1477

360 EPAPPTPKPEATTTPKEPAATTTPKSAF--TPKPEAPATPKKAPATTTPKEPAATT 418

1478 SPPTTTPSPPT--TTTTTTPPTTTPSPPTTTPKASATTLTPPTTTPSPPTTTPPTT 1536

419 TTPKEPAATTTPKEPAATTTPKSAF--TPKPEAPATPKKAPATTTPKEPAATT 478

1537 TTPSPPTTTPPTP--PTSTTLTPPTTTPSPPTTTTTTPPTTTPSPPTTTPSPPTTTT 1595

479-PAPPTTKSAF--TPKPEAPATTTPKSAF--TPKESPTTPKEPAATTTPKEPAATT 536

1596 PPPTTTPSPPTTTTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1655

QY	537	TPKRPATPTKEBAPTTTKKRAPLAKEPARTTPEKALPTPKKLIPRTPEKLAAPTPEEK	596
Db	1656	TSPSPPTTSPSPITTTTTTPPTTSSDITATP--SPTTTS--TMPTPS--PTTTPSS	1707
QY	597	PAPRTPELAPTPPEDEPTPTPEEPAPTTPKAAAPTRKEAPRTTPEAPRTTKEBAPT	656
Db	1708	PITTTTTSSSTTP-SPPRTTMTPTSPPTTTPSPPTTMTTLRPTTSSSLTTLPLPSIT	1766
QY	657	TP-----KETAPTP-----KG-----	668
Db	1767	PRPSPESTPTPTPCVPLCNMTGWLDSGRKNPKPGSDTELIGDYCGHMAANISCAT	1826
QY	669	-----TAPTLK	675
Db	1827	MYPDVLPQLCQTIVVCVSVGLICKNEDOKRGGVIMACLNELINWQCCECVTOPTMM-	1885
QY	676	EPAPTPEKPAKRELAPT--TEKISTSDKP----APL-TPKGAAPTPEKBAPTPK	728
Db	1886	--TTTTENPRTPIITTTTVTPRTPTQSNGLOARPRISITTTVTVPRTPTGTQ	1943
QY	729	EPAPRTPEKGAPTLKEBAPTRPKKPAKELAPTTKGPTST-TSDKPAPTPEKAPRT	787
Db	1944	TPT-TPTTTTTVTVPRTPTGTPTTVLITTTTMTPTPTSTSKSTVAPITTTTTV	2002
QY	788	PKBPAPTPKKP-----APTPETPPTSEVSPTTK-----EPTTIKSPEESPT	835
Db	2003	TATPTPTGTQPTMTPISTTTVTPTPTPTGSTGPTHTNSTAPLAEIATNSNPRESSTP	2062
QY	836	ELSAEPTPKALENSKEKGVPTTTPAATREMTTAKDKTERDLATTPETTAAPKMT	895
Db	2063	QFSRSTSSLTESTLLSLTPRALEMSTAPSRTPAPTTSGHLTSPRSSTTSPGCT	2122
QY	896	KETATTEKTETESKTYATTOVSTTODTPEKITT---LKTTLTAPKYTTTKIIITTT	952
Db	2123	PIRGITT--GSSSAPRPSHVQTTTTSANTPPTPLSPSLIRTTGLRLPYPSVLICVLN	2180
QY	953	ELNNKEF-----TAKPKDRATNSKATTPRKPOKPTK	983
Db	2181	DYTAAGEEVNYGTGDTCFYVNCSLCTLEFYNNSCSTSPPTPSK-STPSPKSPSS	2239
QY	984	APKPKPTSKPKPMVRKPKTTPTRKMTSTMPELNFTSIAGNAMQTTRPQPTNSK	1043
Db	2240	TPSKPTPTGKTPRCPCPFDPFR-----QENETMWLICCFM-AFCIKYNNTVEIV	2285
QY	1044	LVEVNP-----KSEDAGS-----AEGETPHMLL-----R	1067
Db	2286	KVDEPPRMFTCSNGIQPVKVEDPDGCCWHMEDCCYCIGWMD-PIHYVFQDLIYSYGNC	2344
QY	1068	PHYVMEVTPDMVDLPRVNOGIILN	1093
Db	2345	TYLVIEIISPVD-----NFGVIID	2364

RESULT 2
T18262
S-layer protein - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18262
R:Fujino, T.; Beguin, P.; Aubert J.P.
J. Bacteriol. 175, 1891-1899, 1993
A>Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e
A:Reference number: Z18847; MUID:93209931
A:Accession: T18262
A>Status: Preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1664 <NU>
A:Cross-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAAA7841.1

Query Match 13.0%; Score 950; DB 2; Length 1664;
Best Local Similarity 31.0%; Pred. No. 9.5e-33;

cyst germination specific acidic repeat protein precursor - Phytophthora infestans

C:Species: Phytophthora infestans (potato late blight agent)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C/Accession: T31108
R:Goernhardt, B.
submitted to the EMBL Data Library, April 1998
A:Reference number: Z20986
A/Accession: T31108
A>Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1489 <GOE>
C/Cross-references: EMBL:AF061185; NID:g3851513; PID:g3851514; PIDDN:AAC72308.1
C/Genetics:
A/Gene: car90

Query Match 11.8%; Score 867; DB 2; Length 1489;
Best Local Similarity 32.2%; Pred. No. 2,5e+29;
Matches 367; Conservative 52; Mismatches 503; Indels 216; Gaps 51:

Dy 114 PSSKKAAPPSGASQTKSTTKSPK-----PPKK---KTKKVIESEITEVKNKKNR 164
Dy 303 PSSETAPPEGTGYVPREETTAAPSDDTYARRETPYAPAKPDVEETTVYEESTYA 362
Dy 165 TKKKPTPKPVVDAGSGLDNGDFKYT---TRDYSTQHNNVSTSP-KITAKINRP 219
Dy 363 PRKSETNAFLERKHVAHIEKRDDTEVTMAPLEETTYAPLEETTYAPLEETP 422
Dy 220 SLPRNDTSKETSLIVNKCTYETKEETT--TNKQTSDGCKEKTSAKESIEKTSAK 276
Dy 423 YEP----TEETTYPTEETTYAPTEETTYAPTEKTYAPTEETTYAPLEETPYEPTEET 477
Dy 277 DLAPTSKVLAKETPAEAETTTKGALTLTREKPRTTPKBPASTPKREPRTTIKSAPTPK 336
Dy 478 TYAPTEETTYAPT-----EETTVASTBEETTYAPLEETTYAPAETPEPEET-TYAPTEET 533
Dy 337 EPAPPT-TTKSADPTKBPAPT--TTKEBAPPTPKBPAPTTTKBPATTKSAPT----- 387
Dy 534 TYAPEETTYAPLEETTYAPLEETTYAPAEETPYEPTEETTYAPLEET-YAPEETMYA 592
Dy 388 ----PRKEBAPPTPKBPAPTTPKBPAPTTPKREPPTTPKBP-----AP--- 426
Dy 593 PIEETTYAPTEETTYAPAEETPYEPTEETTYAPAEETTYAPTEETTYASTBEETYPTEE 652
Dy 427 TTKEBAPPTPKBP-----APTAPKKAPPTPKBPAPTTPKBPAPT--TTKESPPTPK 478
Dy 653 TYAPAEETPYEPTEETTYAPLEETTYAPLEETTYAPLEETTYAPLEETTYAPAEETPYE 712
Dy 479 P-----APT-TTKSAPT-----TKBP-----APT-TTKSAPT--TPKESPT 512
Dy 713 PREETTYAPTEETTYAPLEETMYAPIETTYGPTPEETTYAPTEKTYAPLEETPYAPTEE 772
Dy 513 TTKEP-----APTTPKBAPPTPKBPAPTPKBPAPT--PRKBAPPTTKKAPATPAKE 564
Dy 773 TTYEPTGETTYAPTEETTYAPLEETTYAPLEETTYAPLEETPYEPTEETTYAPLEETPYE 832
Dy 565 PAPTPPKETAPTPKKLTPTPPEKIAPTPPKBAPTPPEELAPT--TPBEPPTTPPEEP- 621
Dy 833 PTLET--TYTPEETTYAPTEETTYAPLEETTYAPLEETTYAPLEETPYEPTEETTYAPT 890
Dy 622 ----APT-----TPKAARNTPKBPAPTTPKBP-----APTTPKBAPPTPK 660
Dy 891 KETTYAPLEETTYASTEETTYAPLEETTYAPAEETPYEPTEETTYAPTEETTYAPEETP 950
Dy 661 TAPTPPKGAPT--TLKEBAPPTPKBPAPK-ELAPT--TTKEPTSTDSPKAPPTPKGA 715
Dy 951 YAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPLEETTYAPLEETMYAPIETTYAPT 1010
Dy 716 PTTKEBAPPTPKBP-----APTTPKGAPT-----TLKEBAPPTP 751
Dy 1011 EETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETP 1070

[illegible]

RESULT	4
--------	---

hypothetical protein F35A5.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C:Accession: F16251
R:Leimbach, D.
submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid F35A5.
A:Reference number: Z18485
A:Accession: F16251
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1274 <LEI>
A:Cross-references: EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AAB52641.1; GSPDB:GN00
C:Experimental source: strain Bristol N2; clone F35A5
C:Genetics:
A:Gene: CESP:F35A5.1
A:Map position: X
A:Introns: 1272/2

Query Match	11.4%;	Score 835;	DB 2;	Length 1274;
Best Local Similarity	27.5%;	Pred. No. 4.7e-28;		
Matches 310;	Conservative 111;	Mismatches 446;	Indels 260;	Gaps 60

```

OY 95 KCCPDYSCFAEYHNPIRSPSSKAP-----DPGASQIETKTR 135
Db 189 KAAFSKE-----HDPVLPPTPLKNPKAKKPMWEDEVTEELKEPEPATKRVPLAKK 242
OY 136 SP-----KP-PKKKTKKVISEETLEVKDMKKNKTK-KP-----TP-----KPVVDEA 179
Db 243 EPSTSVKPVSDPSTKVKVPKKEPEVPTPIKNPTKKKPMWEDETVVEEKVEKPEPK 302
OY 180 GSGIDNGD-----FKVTPPDTSTTOHKNVSTSPKITTAKPI-NRPSLIPNSDTSKETSLT 234
Db 303 APVLKKNDPAAPAAKAROPSPSKAPKVEBPSPVPTPVKNPKKVKYPMEWEDDEBAEE 362
OY 235 VNKETVETKETETTTNKQSTDGCEKKTSAKETIOSIETSAKDLAPISKVLAPTPRAE- 293
Db 363 YKBPAPPEKKTIVLKRKEPEPSSSTPPSSDSPKKAAPAVKPRDSSPKAPATLQADPPAAOE 422
OY 264 ---TTTKPA-----LTPPKETPTTP-----KEPASTPKPE-----TP 325
Db 423 VPPPVKKPKPVKKYKRPMEWEDDEDEVEEKKOEAAKAKTTPVLKRRKEPAKDTAKPATISKTP 482
OY 326 -TTTISAPLTPPREAPPTTTSKAPTTPKREPAPTTPKEA-----PTT 365
Db 483 ETPEKKDVPYKPRDSSPKVAAKPPDSQAQAP-TPVKNVYKKMRPMEWEDDETADVDKSPTD 541

```

Qy	366	PKREAPTTKEAPPT-----TTSAPTTKEP-----APTPKKAPPTTPE-----PA	409
Db	542	AKTTPSLAKDPAKAKESILKPKADTKAKAKRDRDPSPKVAPTAPEKTTPLAKKEPAGPA	601
Qy	410	PTTPKEPTTPEKEAPPTKEBAPTTPK-EPAPTAKKKAP-----TTKEBAPTTPE	462
Db	602	DSMTKEBESKRPDPSPKKAVPAKPVPTTEVAPAAVKKPEPISKPKDTAKKAPNSPVV	661
Qy	463	PAPTTKEP-----SPTTKEBAPTTTKSAPTTTKBAPTTTKSAPTP-KEPSPTTKEP	517
Db	662	P-PTPVKNPVKKKKPWEDEDDAPAKPVSLPEBEK-TVYLAKKAPITPDSAAADVSGP	719
Qy	518	APTPPK-EBAPPTPKKAP- TTKEBAPTTPKBAPTTTKKAPTA- PKBAPTTPE	572
Db	720	SSKDPLAKKAPKAPKPDSPMKAVPLKAPKI- EVPAVVKKEBPAKSDPSPKAK-	776
Qy	573	TAPTPEKLTPTTP-----EKALPTPEKAPPTPEELAPTTPEEP-----TPTTPEPA	622
Db	777	AEPNSP-VVBPPTPVKNPVKKKPKWEDEDDAAEBVNVNPEBEKTPVLAKTTPKPDPS	834
Qy	623	PTTAKAAPPTPEBAPTTPKBAPTT- PKBAPTTPEPTAPTPPKGAPTTKEBAPTT	681
Db	835	PKKAVPKPSKTKDAPVSVKKKEPVSKPEPSPKAEBSNPVP-----PTPVKNPVKKM	890
Qy	682	PKKAPKELAP- TTKEPSTTSDDKAPPTPKGAPTTKEBAPTTKEBAPTTPKGTAP	740
Db	891	KPWEDDEPTPEVKKPSE- PEKTTPLAK- KEBEKKD- APKVAAPKPDPSPKAVP	945
Qy	741	TTLEKAPPT-----TPKKPAP-----KELAPTTKGTS-----TTSOKP-----AP	777
Db	946	E-KEBPAKVAAPKPHDSLPPKKAIPIPANTOEBAPTPVKNPVKKMKPWEDEDEPEVSA	1003
Qy	778	TTPEPT-----APTTKEBAPTTPKKAPPTPEPPPTTSVSPTTKEPTTIHNSP-	830
Db	1004	EPEKTTPLAKKAKAKRDRD---SPKKAAPVAAP-PDPAIPEV-PPTPVKNPVKKMKP	1058
Qy	831	---DESPPELSA-EP---TPKALENSKPEGV-----PTTKPAT-----	864
Db	1059	EDDEPSPSPVAPPEPEKTTPLAKKAPTKATPDESEAAADVSGTSDKPLSKAPVE	1118
Qy	865	KPEMTTAKOKTTERDLKTTPETT-TAAPKMTKEPATTTTEKTESKITAITTQVTSIT	922
Db	1119	KPKPTTDEKDKLKPSPAKPKPEKAPBEPAKWKMPWMDDPDEPEADFTVAPASKPDTE	1178
Qy	923	QDTPEFTTTLKTTLAPKXVTTTKITTTTEINNKPEGTAKPKDRATNSKATPKPKPT	982
Db	1179	DPADPLG-----GKTKDPK-----LNKAPAKPEK-----PKPKEVS	1213
Qy	983	KAPKPESTTKKPKTMPRVKPK-----TTPTKMTSTMP	1017
Db	1214	KBPPEPTPEPKP-AAPKMKKPMWEDDEDEADFTMPAPAKKPDTEDP	1259

RESULT

I:0020
 desc: nucleotide polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
 N:Alternate names: alpha NAC protein
 C:Species: Mus musculus (house mouse)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
 C:Accession: T30826
 R:Yotov, W.V.; St-Arnaud, R.
 Genes Dev. 10, 1763-1772, 1996
 A:Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle
 A:Reference number: Z20889; MUID:96312450
 A:Accession: T30826
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2187 <Y>
 A:Cross-references: EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB18732.1
 C:genetics:
 A:Gene: Naca
 A:Map position: 10

QY 647 PTPKKEPAPPTPKETAPPTPKGTAPPTLKE---PA-----PTPKK---PAKELAPPTT 695
 Db 918 PAAVSSP-PMTPKSSPPPVVSSPPVVKSSPPAPVSSPPAPVSSPPAPVNLPEV 976
 QY 696 KE---PTSTTSKDPAPPTPKGTAPPTP-KEPAPPTPKETAPPTPKGTAPPTLKEPAPPTT 751
 Db 977 KSSPPPTVSSPPPA---PKSSPPAPMSSPPPEVKSSPPAPVSSPPVSSPPAP 1033
 QY 752 KKAPKELAPPTTKGPTSTSDKPAPI-TPKELAPPTPKAPPTPKKAPPTPEPTPTT 810
 Db 1034 VSSPP---PPVKSPPPPAPVSSPPVKSPPAPVSSPPVKSPPAPVSS-SPPP 1088
 QY 811 TSEVSPPTTKKEPTTHKSPDESPELSAPPTKALENSKEGVPPTTKPATKPEMTT 870
 Db 1089 VKSPPPAPVSSPPPKSPPPAPVSSPPAP-----VKPPSLP-PPAPVSSPPVPT 1141
 QY 871 TAKDKTERDLRTPTPT 867
 Db 1142 PAPPKKEQSLPPPAES 1158
 RESULT 7
 S48478
 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
 N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR019c
 C:Species: Saccharomyces cerevisiae
 C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 12-Nov-1999
 C:Accession: S48478; A26877; B26877; S27281; JG6123
 R:Rowley, K.
 submitted to the EMBL Data Library, October 1994
 A:Reference number: S48478
 A:Accession: S48478
 A:Molecule type: DNA
 A:Residues: 1-1367 <ROW>
 A:Cross-references: GB:Z47047; EMBL:Z38061; NID:9603997; PID:9763364; GSPDB:GN00009; MIF
 R:Yamashita, I.; Nakamura, M.; Fukui, S.
 J. Bacteriol. 169, 2142-2149, 1987
 A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
 A:Reference number: A91831; MUID:87194600
 A:Accession: A26877
 A:Molecule type: DNA
 A:Residues: 1-242 <YAM>
 A:Cross-references: EMBL:M16164; NID:9172522; PIDN:AAA35014.1; PID:9172525
 A:Accession: B26877
 A:Molecule type: DNA
 A:Residues: 762-1331 <YA2>
 A:Cross-references: EMBL:M16165; NID:9172523; PIDN:AAA35015.1; PID:9172526
 R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
 FEBS Lett. 239, 179-184, 1988
 A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
 A:Reference number: S27281; MUID:89031230
 A:Accession: S27281
 A:Molecule type: DNA
 A:Residues: 1-31 <PAR>
 A:Cross-references: EMBL:X13857; NID:94551; PIDN:CAA32069.1; PID:94552
 R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Precorius, I.S.
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
 A:Title: Muc1, a mucin-like protein that is regulated by Mes10, is critical for pseudohy
 A:Reference number: JG6123; MUID:96523237
 A:Accession: JG6123
 A:Molecule type: DNA
 A:Status: nucleic acid sequence not shown
 A:Residues: 1-1367 <LAM>
 A:Cross-references: GB:U30626; NID:91304386; PIDN:AAC49609.1; PID:91304387
 C:Genetics:
 A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
 A:Cross-references: MIPS:YIR019c; SGD:S0001458
 A:Map position: 9R
 C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
 F:1350-1366/Domain: transmembrane #status predicted <TM1>
 F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 10.9%; Score 795.5; DB 1; Length 1367;
 Best Local Similarity 27.4%; Pred. No. 2.3e-26;
 Matches 318; Conservative 111; Mismatches 525; Indels 207; Gaps 51;
 85 DCDACCKKDKD---CCDYDESFCAEVHNPTSPSSKAPPPSGASQNTKSTTKSPKPN 141
 Db 174 DLSTGCNNYDNOGHQSOTDFQFYNNIDCNDNCGSTKSTTSSSES- -STTSS----- 226
 QY 142 KKTKRVIESEETEVNDKNKRTKKKPPVVDAGSLDNGDKVTPPTSTTQH 201
 Db 227 -----TSSSTTSTSSST----- -TTSSSESST 253
 QY 202 KVTSPKI-TTAPINRPSLPNSDTSKETSLTNKETTVERKETTNTKQISTDCKE 260
 Db 254 SSTTAPATPTTSCTKKPPPTTSCTKKPPPHHDPTTCKKTKTTSK-TC- -KK 309
 QY 261 TTSKETSIEKTSKADLPTSKVLAKPTPKAETT--TKGPAITTPKE-----PTPTT 312
 Db 310 TTPVPTPS- -STTESSAPV-----PTSSSTTESSAPVSTTESSAPVPTPSSS 362
 QY 313 KEPASTTPKEPTPTTKSAP---TPKKEPAPTTKSAPTTKKEPAPTTKKEPAPTTKE 368
 Db 363 TTESSAPVNTSSTESSAPVNTSSTESSAPVPTPSSSTTESSAPVNTS- -TTSS 418
 QY 369 PAPTTPKEPAPTTTKSAPTTKKEPAPTTKPKAPPTPKKAPPTPKKPTT- -TPKKEPA 425
 Db 419 SAPVTS- -STTESSAPVNT- -STTESSAPVNTSSTESSAPVPTPSSSTTESSA 471
 QY 426 PTPKKEPAPTTKKEPAPTKKAPKAPTTKKEPAP- -TPKKEPAPTTKKEPAPTTKKEPAP 481
 Db 472 PVT- -STTESSAP- -VPTPSSSTTESSAPVNTSSTTESSAPVPTPS- -STTESSAP 526
 QY 482 TTKKSAPTTKKEPAP- -TTKSAPTTKKEPAPTTKKEPAP- -TPKKEPAP- -TPK 532
 Db 527 APPTSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSS 586
 QY 533 PAPTTPKEPAPTTKKEPAPTTKPKAPKAPTTKKEPAPTTKKEPAPTTKKEPAPTT 592
 Db 587 STTESSAPVPT- -PSSSTTESSAPVPTPSSSTTESSAPVPT- -STTESSAP- 637
 QY 593 TPKKPAPTTPPELAPT-TPKEPTTPPEEPAP- -TPKKAAPNTPKKEPAPTTKKEPA 646
 Db 638 VPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPT- -STTESSA 693
 QY 647 PTPKKEPAPTTKKEAPPT-TPKGTAPTTKKEAPTTKPKAPKELAPTTTKEPTSTGK 705
 Db 694 PVT- -STTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPS- -STTESS 748
 QY 706 PAPTTPKGTAPTTKKEPAP- -TPKKEPAPPT-TPKGTAPTTKKEPAPTTKPKAPKELAP 761
 Db 749 SAPVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAP 808
 QY 762 TTKKGPSTTSKDPAPPTPKETAPTTKKEPAP- -TPKKEPAPTTPEPTPTTSKVT 816
 Db 809 VPTPSSSNTTSSAPSTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPT 868
 QY 817 PT- -TKKEPTTHKSPDESPELSAEPPTKALENSKEGVPVPTTKPATKPEKTTT 871
 Db 869 PSSSSNTTSSAPSI- -PSSSTTESFSTGT-TPVSSSKVPSQSTETSSTETTTIVP 924
 QY 872 AKDKTERDLRTPTPTTAAKMTKEATTEKTESKIRAT- -TQVNSTTQODTTPK 929
 Db 925 TKTTTSTTPTSTTTTITVTCSTGNSAGETTSKSPKIVTTPPTTITVSTTSTTTIT 984
 QY 930 ITLTKTTTLAPKVT- -TKKTTTT-ELMNKDEETAKPKDRATNSKATPKPKPK- -981
 Db 985 TTVCSGNTNAGETTSKSKITTTTIVPCSTSPSEA- -SESTTSTTPTTIVT 1036
 QY 982 -----TKAPKFTSTKKKKTMDRVKPKTTPTPRMTSTMPELNPTSRIAE 1028
 Db 1037 VSTTVVTVTEYSTTKGGETTTTFVTKNIPPTVLTITLAPPT- -SVTVVNFPTPT-ITTT 1093

QY 1029 MLOTTTRPNOTPNSKLVEVNPSEDAAGETPHMLRPHVEMPEVTPMDLYLPRVPO 1088
 DB 1094 VCSIGF-----NSAGETSSGSPKVTVTTPCSTGTGETTATTLVTAATV 1140
 QY 1089 GIITINMLSDENICNGKPPDGLITLRLNGTLVAFRGHYEMMLSPSPSPSARITEWGI 1148
 DB 1141 TTVVTESSSTGNTSA-GKTTTGYTKSVPTT-----YVTTLAPASAPVPAATN-----AV 1188
 QY 1149 PSPIDVTFTRC-----NCEGKT 1165
 DB 1189 PTTITT--TECSAATNAGET 1207

RESULT 8
 T18535
 high molecular mass nuclear antigen - chicken (fragment)
 C:Species: Gallus gallus (chicken)
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T18535
 R:Shimada, K.; Harata, M.; Mizuno, S.
 J. Cell Sci. 110, 3031-3041, 1997
 A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, hNNA, of chick
 A:Reference number: Z18955; M0ID:9803440
 A:Accession: T18535
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1151 <SH1>
 A:Cross-references: EMBL:D88440; NID:d117138; PID:d1023045; PIDN:BA24137.1

Query Match 9.6%; Score 704.5; DB 2; Length 1151;
 Best Local Similarity 25.7%; Pred. No. 1.2e-22;
 Matches 304; Conservative 104; Mismatches 445; Indels 331; Gaps 59;
 QY 105 AEVHNPTSPSSSK-----APPSG-----ASOTIKSTTKRSPKP 140
 DB 39 AELH-PAAPQPPKWPPIGAPPPGTEPTPSPKPTDGADAAKASAEITSPSPSPSP 97
 QY 141 NKKKTKVIESEITEVKDKKKRKKRTPK-----PVVDEAGSGDNGDFKVTTP 193
 DB 98 DGPKAPSGAGEA-----AGTPPSQGPAGTPPSQGAAGAPKGDGTAPQS 143
 QY 194 DNSTGTHNKVST--PKITTAKPINRPSLPPNSDTSKETSLTVNKETVETKETTNNK 251
 DB 144 GTRSGADGKPAQDVPKATTA-ATEAR-----ASAAPVVPKATATATVTAASQS 194
 QY 252 --QTSIDGKEKTTSAKETTSIEKTSKDLAPISKVLAKPTPKAETTTKGPALTTPKEPT 309
 DB 195 APKAAADAAAVTAA--SOSAPKATV-EVKPAAAVAAKAKAVTAAVAAAFKATAEAKPAP 250
 QY 310 -TPPKREPASTPKP-----TPTIKSAPTTPKEPAPT-----TKSAPTTPKEP 353
 DB 251 VISPPTIPCSSAEKPLTAASPTASKA--TAEAKVPATATSLMAKTKYTAEKPPSPSPSVK 308
 QY 354 APTTK-----EPAPTTPKEPAPTTPKEPAPTTPKSAPTT-- 388
 DB 309 ATDTKAVATATAPKAGPDYKPAVAVCAEAKPAPPPQQAAPAAAAAGTGLKPAATAP 368
 QY 389 ----PREPA-----PTTPKKRAPTTPKEPAPTTPKEPTT-----PKRPAATTEPAP 433
 DB 369 PHGSPRANSHVTVPNPVPRAAATVP--TAGAVPKASGTTPAAPQOPVPV--KAAP 423
 QY 434 TTPKEPAPTAPKAPAPTTPKEPAPTTPKEP-----APTTPKEPSTTPKEPAPTTPKSAPTT 490
 DB 424 VTFPSPOQAVPRAATAAA-----APVTPOQVYTAATTTNATPPQPIPKAATTTTATPVT 479
 QY 491 TKEPAP--TTTKSAPTTPKEPSPPTTKEP-----APTTPKEPAPTTPKAPAPTTPKEPAP 543
 DB 480 PQQPIPKAGIDAAPPAVPKAPSDGRAATPGVNAATDPQKPPPTQSVSPSAVTEPKPQ 539
 QY 544 -----TTKEPAPT--TTKKPAPTAPK-----EPAPTTPK--ETAPTTPPKLTP 583

DB 540 RAAPPSNEATPAVPSPSPNLKSPPLPTIPKPVLMALTPQPVTAQVNTOLAATKPSPIVP 599
 QY 584 TTEPKLAPTTPKEPAPTTPPELA-----PTTP-----EPTPTTP--EPA 622
 DB 600 KASPK-ALMTPPPPPPGLPALAAAKLLGLPSSPVASAMAKVTPRLPSPVPMASPA 658
 QY 623 PTPP-----KAAAPNTPK-----PAPTTPKEPAPTTPKEPAPTTPKEPAP 663
 DB 659 SLGPDAARVALATNNAASPGAKPEAGNGNTLAMPKANQOMAPIGAAGAAQTAPMGAH 718
 QY 664 T-----TRKTAPT-----TLKEPAPTTPKAPAKEL 690
 DB 719 THVSPMGAGATQMSPTGANTHMSPIGAGATQMSPMGANQOMSPMGATTTQMSMGA 778
 QY 691 APTTPKEP-----TSTSDKPAPTTPKG--TAPTTPKEPAPTTPKEPAPTTPKGTAP 740
 DB 779 AATTOPSPKMAAATQVATISAGNTMOVSGAATTPOTPSVGAATTP-QSPM---GAA 833
 QY 741 TTKEP--APTTPKAPKELAPTTPKG--TSTSDKP-APTTPKET--APTTPKE 790
 DB 834 TTLMSPMGAATTPQ---PSPMGAVTTQPPPMATNTTQPPPMASPTQSPMGAAATTTQS 890
 QY 791 P--APTTPKAP-APTTPETPP-----PTSEVSTPTTPKEPTTIHKSPPDESTPEL-- 837
 DB 891 PPMGATTTQSPMGASTPQAPPTVAGSPTPPPPIPSPTQTSPOPMKSPPPDPKAPS 950
 QY 838 SAPPTPKALENSPKKEGVPTTKTPAATKREMTTAKDTERLRLTPETTAAPKMT-K 896
 DB 951 AAGQTSFPAHVANASPEV-TAVSPA--PIGVTEASPADGARISPEPTAATGPKASPA 1006
 QY 897 ETATTEKTESKITAATTOVSTTQDTTPFTITTLKTTLAPKYVTTTKKTTTTEIM 956
 DB 1007 ATADVTEAARD-VTAAATVPA-----EAP-----TKAKRSSSSSSSSSSSS 1051
 QY 957 KPEETAKPKDRATNSKATTPPKPKPTKAPKPTSTKPKMPVRYKPKTTPPKMTSTM 1016
 DB 1052 SSSSSSSSDSSSSSSSESNPASPA-----PAVGDCQOQMTGAAQSV 1096
 QY 1017 PELNPTSRIEAMLOTTTRPNOTPNSKLVEVNPSEDAAGAGE 1060
 DB 1097 P-----VTEAAVQ-----EAAAAAAGAEERB 1119

RESULT 9
 T25697
 hypothetical protein F16F9.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T25697
 R:Fulton, B.
 submitted to the EMBL Data Library, August 1996
 A:Description: The sequence of C. elegans cosmid F16F9.
 A:Reference number: Z20071
 A:Accession: T25697
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1229 <Ful>
 A:Cross-references: EMBL:U67956; PIDN:AB07691.1; GSPDB:GN00026; CESP:F16F9.2
 A:Experimental source: strain Bristol N2; clone F16F9
 C:Genetics:
 A:Gene: CESP:F16F9.2
 A:Map position: X
 A:Introns: 35/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 9.2%; Score 674; DB 2; Length 1229;
 Best Local Similarity 27.8%; Pred. No. 2.4e-21;
 Matches 328; Conservative 89; Mismatches 384; Indels 380; Gaps 56;
 QY 73 RCFESFERGECDCDAQCKYDKCCPDYSEFCAEVHNPTSP-----PSSKKAAPPS 123
 DB 34 RDKRYKHKHKNTCTGSC-----KCVDPASNPFDVSTTSSINNDVNDIGPSGDSNPTGS 89

```

QY 124 GASOTIKSTKRSRPPKPKKKKVKISEETVEKONKNNRKKPKPPVVDAGSGL 183
DB 90 SMFOELAIIVGGO----TVKSEHNIDSSVEY-----EKKVT----- 121
QY 184 DNGDFKVTTPDSTTGO---HNKYSTSPKITTAA-KPINDRSLPENSNDT-----SKETS 232
DB 122 -----TSDASTNAPITGKDSSTPEIITGIIVINSKSSSVYDMSTPRSTLSPTTE 174
QY 233 LYNNKETTVEKETTNNKOTSDGKFKTTSKAKTOSTEKTSAOIDLAPTSKVLAKPPPKA 292
DB 175 LILSPLELVSTDSSTST-EOTSPPONTIELASPMETNTTATTTTSVPSVSTLAS---ED 230
QY 293 ETTTKGALITPKPEPTTPTEKPEASTPKPEPTTIKSAPTTPKEPAPTTPKSAPTPKPE 352
DB 231 ETTVTLAEST-----TVIAEKVSTTEPTTT-----AESITTKKSTT----- 268
QY 353 PAPPTTKPEAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTP---KKRAPPTPKPE 408
DB 269 KAPATPEEPPTTTEE---VTTTEAETSTTSSSETSEK---PTTPLIDNKIAGPATGK-- 321
QY 409 APPTPKPEP-PTTPPKPEAPTTPKEPAPTTPKEPAPTTPKAPATPKAPAPTTPKEPAPT 467
DB 322 -PETTHFVGTTPN--FDPAETPEVAKSEDKMFLSKTATETTOOTTEVT-DGPEKET 377
QY 468 TRE-----PSPPTPKPEAPTTPKSAAP-----TTTKPEAPT---TTKSAPTTPKEPSP 511
DB 378 TKNVSEIPIITTVPLVETSTSTFASKESDGFHTTLKLTATNDSDESATITVKKPNEE 437
QY 512 TTTKE---PAPT-----TPK-----EPAPTTPKAPAP----- 535
DB 438 TTTKSHVVPKPKKGVKVPKLELSPEDETEIT-KAPHGKLEKKTYHFVLSDNFARY 496
QY 536 TTPKEPAP-----PAPTTPKEPAP-----TTTKKAPAPA--PKPEAPT---PK 571
DB 497 SEAKENDVYHMDYNNKREKPEPTTTEESSTTEVEVITTEPANTGNPTTENPTTTOPT 556
QY 572 ETAPPTPKKILPTTPPEKLA---PTTPKEPAPTPEELAPPTPEPTTPTEEPAPPTPKA 628
DB 557 STASSTTALPFTTEQVUTTEEPTEAKSTATO---KPTTQOSVST---EKSTTKKA 609
QY 629 AANTREPEAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKGTAPPTTLKEPAPTTPKAPAK 688
DB 610 ---STTEE--PTTDEPTT---ESSTTKATTPPELSTTSEETTELKITE----- 655
QY 689 ELAPPTTKPEPTTSDKAPAP---TPPKGTAPPTPKPEAPTTPKEPAPTTPKGTAPPTTLK 744
DB 656 --GSTTTEEPPTTALPFAEASTGIITDEETSTTSTPEITSTKE--IVESATITOTSVS 711
QY 745 EPAPTTP-----KKRAPKEAPPTTKGP----- 767
DB 712 VVESPTRQLEPRKWKALVKNFKHNLVLEKKEKRLKKEKESSTGSDSSSETTVVAENID 771
QY 768 --ISTSDKAPPTPKKATPTTPKEPAPTTPKPKAPPTTPPE--TPPTTSE-VSIPPTTKE 822
DB 772 EVTTTEKKEKVQPTTTEKSTTOEETTTTTEKTTSTTEKTPPTSESATTEPTTSE 831
QY 823 PTTIHKSPDSSTPELSAEPKPKALENSPKRPGVPTTKTPATKREMTTTLKDKTTEDLR 882
DB 832 PST-----TEST-----TVDTSSATTEESSTAETTTTSAE-- 862
QY 883 TPETTTA-----APMTKEPTATTE 903
DB 863 -TSTTTTSESAAPIITGESPPENTALQSSQKSEENESSAERKRGARNDVVRKKNHTTVRAE 921
QY 904 KTESKTKTATTPQ-VIST---TTQDTPPKITTLKTTTLAPKYT---TKKITTTEIIM 956
DB 922 TTSAAVASTTTEPTTTEKSTLETTETPEATILNEVTGAFATVGADVDETTITTELLS 981
QY 957 KPEETAKPKRATNSKATTPKPO-----KPKAKPKPTSTYKPKKPMR 999
DB 982 K-----INNTQISQKPPDISKTALSSISGLISGTFKAPMAPTI----- 1022
QY 1000 VRKPKTTTPRKMTSTMPLEIN-----PTSRIDAA 1028

```

```

DB 1023 -----HTTDAFAVTATFASLNDGSDKKIIDEAOPTDELIRRA 1059

RESULT 10
A35175
mucin 1 precursor, repetitive splice form A [validated] - human
N/Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; epista
ncretatic mucin; polymorphic epithelial mucin (PEM)
N/Contents: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor
C/Species: Homo sapiens (man)
C/Date: 20-Apr-2000 #sequence-revision 20-Apr-2000 #text-change 02-Jun-2000
C/Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218;
R/Biot. Chem. 265, 5573-5578, 1990
J. Biot. Chem. 265, 5573-5578, 1990
A>Title: Epistatin, a carcinoma-associated mucin, is generated by a polymorphic gene
A/Reference number: A35175; MUID:9020794
A/Accession: A35175
A/Molecule type: mRNA
A/Residues: 1-952,1033-1344 <LIG1>
A/Cross-references: GB:M32738; GB:J05288; NID:9182121; PIDN:AAA35804.1; PID:9182124;
A/Experimental source: splice form A
A/Note: GenBank entries HOMEPIST1 and HOMEPIST2 present only the amino- and carboxyl
A/Accession: B35175
A/Molecule type: mRNA
A/Residues: 1-19,29-952,1033-1344 <LIG2>
A/Cross-references: GB:M32739; GB:J05288; NID:9182126; PIDN:AAA35806.1; PID:9182129;
A/Experimental source: splice form B
A/Note: GenBank entries HOMEPIST1 and HOMEPIST2 present only the amino- and carboxyl
R/Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Dunig, T.; Peat, N.; Burc
J. Biot. Chem. 265, 15286-15293, 1990
A>Title: Molecular cloning and expression of human tumor-associated polymorphic epilth
A/Reference number: A35886; MUID:90368715
A/Accession: A35886
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-19,29-992,1033-1344 <GEN>
A/Cross-references: GB:J05581; NID:9188869; PIDN:AAA59876.1; PID:9188870
A/Note: GenBank entry HUMPMNCAB includes one copy of the tandemly repeated sequence
R/Jian, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
J. Biol. Chem. 265, 15294-15299, 1990
A>Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
A/Reference number: A35887; MUID:90368716
A/Accession: A35887
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
A/Cross-references: GB:J05582; NID:9189598; PIDN:AAA60019.1; PID:9189599
A/Note: GenBank entry HUMPMNCAB contains four fewer copies of the tandemly repeated se
R/Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky,
Eur. J. Biochem. 189, 463-473, 1990
A>Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may gen
A/Reference number: S10571; MUID:90276413
A/Accession: S10572
A/Molecule type: mRNA
A/Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>
A/Cross-references: EMBL:X52229; NID:937053
R/Wreschner, D.H.
submitted to the EMBL Data Library, March 1990
A/Reference number: S40293
A/Accession: S40293
A/Molecule type: mRNA
A/Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <
A/Cross-references: EMBL:X52229; NID:937053; PIDN:CAA36478.1; PID:937054
R/Abe, M.; Sidiqul, J.; Kufo, D.
Biochem. Biophys. Res. Commun. 165, 644-649, 1989
A>Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associat
A/Reference number: A36735; MUID:90088473
A/Accession: A36735
A/Molecule type: mRNA
A/Residues: 1-142,'O',144-162,'O',164-168 <ABE>
A/Cross-references: EMBL:M1823; NID:9181542; PIDN:AAA35757.1; PID:9181543
R/Masuzawa, Y.; Miyachi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu

```


Db 2399 -SEDDLTGSGTENIPASSKQITISSTPTDITLASEEPTKSTMSDLSITTSNVLSSESS 2457
Qy 624 TTPAAAPNPKKEPAPTT-----PKKEPAPTT-----TPKEAP-----TT 657
Db 2458 TTPSSSS-----KSVSSSTEGISVIVSTERSKVEPSTISSVLEBDLTKTTPILLEBT 2513
Qy 658 PKETAPTPKGAAPTTTKEPAPTTPKKPAKELAPTTTKEPTSTSDKA-----P 708
Db 2514 ASETSESLTSDSLVSRINHELTSSENVKRESSTTSSSSKPSQEPAGILSTVVP 2573
Qy 709 T-----TP--KGTAP--TTPK-----EPAPTTKEPAPTTPKGTAP 740
Db 2574 TSSVSLITASEIEAITGNTPEKGRPTITSPKSLVKSTISPTVTSSESESESKRTTSS 2633
Qy 741 TLLKEPAPTTPKKPAKELAPTTTKGPT--STSDKPAAPTTPKKEPAPTTPKK 799
Db 2634 TTVSTTPTETTESLILTAAPSKPTESSTESSEAPTTPAKTSKPSNVSTSKST 2693
Qy 800 APPTPEPPTTSEVSTPTTKEPTTIHKSPEDESTP--EISAEPPTKALE--NSPKPEGV 855
Db 2694 ENVETSTSGSGSLESSTMSSTSS-----SEPETNAPAVTVSSSEASSTTLENSSTSP-- 2745
Qy 856 PTTKTPAPTK-----PKMTTANOKTTERDLRTTPETTTAAKPKTKETATTEKTESKI 910
Db 2746 --TSSEASVLSLFPESITSEAVTVSSR-----APAEITWSSESHREISTVSESEPEI 2799
Qy 911 TATTT-----QVSTTQDQTP--FKIT-----LKTTLAPKVTITTKTI 949
Db 2800 PLSTTVSPNVVATSSISEEPISSVTSSTPRVRLITGTPDDLVTVVPSHGNNRONT 2859
Qy 950 TTEIMNKPPEETAKPKDRATNSKATTPKPKPTKAPKPTSTKKPTMPVRKPK--TT 1006
Db 2860 TASSV--PNSNSPIILPSES--LTTPQPPPTTTTAKPATTSKRRPSPSIQPAEMFTT 2915
Qy 1007 PTPKMTSTMPELN-----PTSKRAEMLOTTT 1034
Db 2916 PAP-----PPSNGXGGEETNOEEQVTSITT 2942

RESULT 13
T45025
muscn MUC5B, tracheobronchial [imported] - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T45025
R:Dessey, J. L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J. P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A:Title: Human muscn gene MUC5B, the 10.7 kb large central exon encodes various alternat
A:Reference number: Z22899; MUID:97166151
A:Accession: T45025
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3570 <DES>
A:cross-references: EMBL:Z72496; MID:g1834502; PIDN:CAA96577.1; PID:g1834503
A:Experimental source: Placenta
C:Genetics:
A:Gene: MUC5B

Query Match 9.1%; Score 663; DB 2: Length 3570;
Best Local Similarity 26.1%; Pred. No. 1.9e-20;
Matches 322; Conservative 92; Mismatches 489; Indels 330; Gaps 54;

Qy 54 MECCDPFKRYCTAELSCGRCFESFERGCDCAOCKKYDKCCPYE--SFCAEV--HNP 110
Db 2279 VESLDEGLYC-----RNR-----QYGR--KICFNEIIVFCNIGHCP 2317
Qy 111 TSPSSSKAPPSGASQT---IKSTTKRSPKPKPKTKTKVISEEITTEVKNKKNTKK 167
Db 2318 STPATSTTAPSSSTPGTWTILTKLT-----TATTESTGTATPSS----- 2359
Qy 168 KPPPKPPVVDASGLDNGDFKVTTPDTSTTQINKVSTSKITTK-----PINRP 219

Db 2360 --TQGP-----AG-----TPHVSTT-----ATTPVTSKATPPSSPGTATAP 2397
Qy 220 SLPPNSDTSKETSGLVNEKETTETVET---KETTNNKQSTDGKEKTSKAKQSIKTS 274
Db 2398 ALRSTATTPTATSPATLIPSSLGTTWRLSQTTTPMATWSAPSSSPERVHTSTVLTT 2457
Qy 275 AKDLAPTSKLAR--TPKMETTKGPAULT---PKERTPTPKRPASTTKERTPTT 327
Db 2458 ATTTGATGSV--APSSSTPGTAHTTKVPTTTTGTGTVTPSSSPGATRPVWISTTTPTT 2516
Qy 328 IKSAPTTPKKEPAPTTTASAPTTKEPAPTT--TKKEPAPTTKEPAPTTTKEPAPTTTGSAP 386
Db 2517 SGSTVTPSSIPGTHHTPTVLTTTQPVATSGMATPSSSTQTSPPSLITATITATGS 2576
Qy 387 TT--PKKEPAPTTPKK--APTTKEPAPTT---TPKEPTPTPKKEPAPTTKEPAPTTKEPA 440
Db 2577 TTNPSSTPGTTPPIPELTTLTATTPAATSSVTPSSALGTHHTPPVMT--TATTHGRSL 2634
Qy 441 PTPAP-----KRPAPTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTKEPA 480
Db 2635 PSSPHVYRTAMTSATSGTLTHTTEPSTGISHTPAAT---GTTTSTPALSSP--HPS 2689
Qy 481 PTTKSAPTT--TKKEPAPTTTASAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTPK 539
Db 2690 SRTTESPPSGTTPPTGHTTATSRATATPSSKRTSLSLBSQPTSAPIIT---VVTTCG 2745
Qy 540 EPAPTTK-----EPAPTT-----TTKKP----- 557
Db 2746 EPQCAMSEMDLSYPMGPGSGDFDTYSNIRAGAVCEQPLGECRATQAPRLGELG 2805
Qy 558 -----APPAKEPAPTTPKETAPT-- 577
Db 2806 QVVECSIDPGLVGNRNOVKFKMCFNYELRVCCNCHGCPSPATSTAMPSSPTGCTW 2865
Qy 578 -PKKLPTTPEKLAPTTPEKAPAPTTPEELAPTTPEEPPTTP--EEAPPTPKAAAPTPK 635
Db 2866 ILTELT---ATTASGSTATPSSPTGTAAPPKVLVSPAATPTATSKATSSSPR 2920
Qy 636 EPAPTTKEPAPTTPKRAPATTTKETAPTTPKGAPTTLKEPAPTTKKAPELAPTT 695
Db 2921 -TATTLPLVLTAKSTATSVTP-----IPSSLTGTLRQTTTPVATMSTIHPSST 2973
Qy 696 KEPTSTSDKRAPATTPKGAAPTTPKKEPAP-----TPKKEPAPTTPKGATPTTKEPAPTT 750
Db 2974 PETHHSTVLTKATPATSTGSPSTGCTWILPFLTAATTTAGTGPTATPSSPTGT 3033
Qy 751 PKRAPKELAPTTT--KGPSTSDKRAPT--TPKETAPTTKEPAPTTPKRAPPTPE 805
Db 3034 TWILTELTATTTATGSTATLSPGCTWILTEPSTATVAPPGSTATVASTQTATG 3093
Qy 806 TPPTTSEVSTPTTKEPTTIHKSPEDEST--PEL--SAEPPTKALENPKPEGVPT----- 857
Db 3094 TPHVSTT--ATTPVTSKATPSSSPGATAPLALRSTATTPTATSPF---AIPSSSLGT 3148
Qy 858 -----TKTPAATKPEMTTAKDKT--TERDLRTTPEIT-----TAAPKTKETATTT 903
Db 3149 TWTRLSQTTTPATMSTATSPSTPEVHTSTVLTTATTTGATGASVATPSSSTGATATTK 3208
Qy 904 --KTESKITAT-----TVOVSTTQDQTT-----PKIT 931
Db 3209 VPTTTTGTATPSSSPGATLTPVWISTTTTPTTPTSSGTVPPSSIPGTHHTARVL 3268
Qy 932 TLKTTTLA-----PKVTTKKTTITTELMNPELV--AKPDORATNSK 972
Db 3269 TTTTITVATSGMATPSSSTGTPSLTATTTATTTAGSTVTPSSPTGTPPVLVTSM 3328
Qy 973 ATPPKDQPTKAPKPTSTKPKTMPVRKPKPTTPPKKATSTMPELNPTSRIAEALQ 1032
Db 3329 ATT-----PAATSKATSSSSPRTATTL--PVLSTATKSTATSPPIPS-----TLMTT 3377
Qy 1033 TTRPNC--TPNSKLVEYNPKSEDAAGAGETTPH 1063

Db 3378 WTVPAQTTPMSMTSTHTSSTP-----ETTH 3404

RESULT 14

151618

nucleolar phosphoprotein - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999

C:Accession: 151618; S57737

R:Cell: 108, 3339-3347, 1995

J:Cell: 108, 3339-3347, 1995

A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xnoppl80

A:Reference number: 151618; MID:96019267

A:Accession: 151618

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-990 <CAI>

A:Cross-references: EMBL:X88927; NID:g895920; PIDN:CAA61368.1; PID:g895921

C:Gene: xnoppl80

C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein

C:Keywords: phosphoprotein

Query Match 8.8%; Score 646; DB 2; Length 990;

Best Local Similarity 25.8%; Pred. No. 2.9e-20;

Matches 260; Conservative 134; Mismatches 389; Indels 226; Gaps 46;

Qy 112 SPSSKAPPPSGASQTIKSTTKRPPKPKTKVIESEITEVKNKNTKTKPT-170

Db 57 SPDAKKRRPPANGI-----PKKSAKESSESDSEDEPPAKKRAQAG 102

Qy 171 PKPVYDEAGSGLDNGPFKTTTPD--ISTQHNKVSTSPKTTAKPINRPSLPPNDS 228

Db 103 GKRRVY-----KAVQPKAKSSSESDSEDEETKPKAPRAQTPK---146

Qy 229 KETSLVNNKETTVEKETTNNKSTDGKKTSAKETOSIEKTSKADLAFTSKVLAKP 288

Db 147 -----VAAVTTPQKKKS-----SSSESSSEDEAKKKQPVAKV---182

Qy 289 TPRAETTKGPAITTPKEPTTPKE---PASTTPKEPTTPKSAPTTPKEPAITTK 344

Db 183 PPKQAVVKAAGIASNNGTAOSSSESDSDSPAKKTAATKTPPT--KPATPAKQAKTA 239

Qy 345 SAPTPKEPAITTPKEPAITTPKEPAITTP-----TKSAPTPKEPAITTP---395

Db 240 GKSSSKRESDSDSEDEOKTAKSPKPDVYSAVPPTSVSKKTLSDPGTKAKESSDSS 299

Qy 396 -----TPKKAPITTPKEPAITTP---KEPTTPKEPAITTPKEPAITTPKEPAITAP 444

Db 300 DSDSEEQPAKKAIVAKAASAPKPLAKAETSTDESDSSSEDEKSSVKKLGVAAP 359

Qy 445 KKPAITTPKEPAITTP---KEPAITTP-----TTTKEPSPTTPKEPAITTP 483

Db 360 KK-APAAP--DAKSTVPAAKKASAPAKKASSSDSSSNEETTTTAKAKTTPAKSAATP 416

Qy 484 TKSAPITTPKEPAITTPKSAPITTPKEPSPTTPKEPAITTPKEPAITTP-----PTTPKKAPITTP 539

Db 417 TSKPTTPGK--ATPTSKTPAKPPTPKTSTAKKSSSDSSSDSEDETTTKPAKKTTPA 473

Qy 540 EPAPT--TPKEPAITTPKPAITAPKEPA--PTTPKEPAITTP-----PTTPKPLTPT 585

Db 474 KSAATPTPSKTP---TNSKATPTSKTPAKPPTPKTSAAKSSSDSSSDSSDKKTPA- 529

Qy 586 PEKIAPTTPKEPAITTPPEELAPITTPPEPTTPPEPAITTP--KAAPNTPEPAITTP---640

Db 530 -KRAAKTTPAKPA-----AKTTPAKPA-----AKTTPAKPAKSTGKQVPTKES 574

Qy 641 TPKEPAITTPKEPAITTPKEPAITTP--KGTAFTLKEPAITTPKKPAKELAPITTPKEPT 699

Db 575 SSSSSSSSSSEDEKSSKPAVKTTPGKAYS-----KPVASKPAK-----KASS 621

Qy 700 STSDKAPITTPKGTAPITTPKEPAITTP--PKPAITTPKGTAPITTPKEPAITTPKKPAK 757

Db 622 SSDSDSEETTTTKTPLTLSPAIVTLPPKKAKSSSDSSSDSDSEK---TPAKPAK 678

Qy 758 ELAPITTPKGP-----ISTSDKAPITTPKETAP--TPKEPAITTPK--PAPITPET 806

Db 679 SAPVNTKAPAKONKASKASCSDSSSEEGKSKOPTGSKPAKAPAKNVAANKKD 738

Qy 807 PPTTSVSTPTTTPKE--PTIHKSPDESTPELSAEPKPLENSPKPEPGVPTTKPAATK 865

Db 739 PSSSSSDSSGDDKQKPKAAAKDKQCAKAKAPTPKKAASSSE---DSSDEDVSK 795

Qy 866 PEMITTAKDITTRBDITTPPETTTAAAPKMTKETATTEKTESKITATTTQVSTTTQDT 925

Db 796 AKTINTAVSKSPV---TPKAVPAAKKESSESDSEDEKQGGKNTSTKIANTST---847

Qy 926 -TPFKITTKTTTTLAPKYTTTK--KTITTTIMNKPEETA-KPRDRATNS---KATTPKQ 979

Db 848 -----PKMAAECSSESSSEDEKANKTSGKRRKRESGAECAYVPE- 891

Qy 980 KPTKAPKPTSTKPKKPMVPRKPKTTPP--RKMTSTMPLELNTPTSTIAE 1027

Db 892 -----NKKLKASPTTPKVNKKELKNTPFRRVEDEINP--RMAD 932

RESULT 15

extensin class 1 precursor - cowpea

C:Species: Vigna unguiculata (cowpea)

C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000

C:Accession: T11622; S54155

R:Aseniljevic-Maksimovic, I.; Broughton, W.J.; Krause, A.

Mol. Plant Microbe Interact. 10, 95-101, 1997

A:Title: Rhizobia modulate root-hair-specific expression of extensin genes.

A:Reference number: 217301; MID:9715574

A:Accession: T11622

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-489 <ARS>

A:Cross-references: EMBL:X91836; NID:g1015936; PIDN:CAA62943.1; PID:g1015937

A:Experimental source: sub-species Red calcona

R:Aseniljevic-Maksimovic, I.; Broughton, W.J.; Krause, A.

Submitted to the EMBL Data Library, April 1995

A:Description: A class of root-hair specific extensins involved in rhizobium/legume i

A:Reference number: S54155

A:Accession: S54155

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 326-489 <AR2>

A:Cross-references: EMBL:X86030; NID:g791149; PID:g791150

C:Gene: Ext26

C:Superfamily: hydroxyproline-rich glycoprotein

C:Keywords: glycoprotein; hydroxyproline

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-489/Product: extensin class 1 #status predicted <MAN>

Query Match 8.6%; Score 633; DB 2; Length 489;

Best Local Similarity 31.2%; Pred. No. 5.2e-20;

Matches 149; Conservative 33; Mismatches 252; Indels 44; Gaps 7;

Qy 335 PKEPAPITTPKSAPITTPKEPAITTPKKEPAITTPKKEPAITTPKKSAPITTPKEBPAP 394

Db 39 PKQTPPYVYNAAPRYVYKSP-----PSPSP-----PPRYVYKTPRYVYKSP 83

Qy 395 TTPKRAPITTPKEPAITTPKEPTTPKKEPAITTPKKEPAITTP---KEPAITAPKRAPITTP 451

Db 84 PSPSPPPPYVYKSPSPSPPPPYVYKSPSPS--PSPPPPYVYKSPSPSPSPPPPYV 141

Qy 452 PKEPAPITTPKEPAITTPKESPTTPKKEPAITTPKKSAPITTPKKEPAITTPKKSAPITTPKEPS 511

Db 142 YKSPSPSPSPPPPYVYKSPSPSPPPPYVYKSP-----PSPSPSPPP 187

```

QY 512 TTKEPAPTPEKAPPTPKKPAFTTKEPAPTKEPAPTTKKPAJAPKEPAPTTPK 571
Db 188 yyykSppSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 247
QY 572 ETAPTPEKKLPTPEKLAFTTPEKAPTEPEELAPTTPEEPPTPEEPAPTTPKAAP 631
Db 248 PPPPYYYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 307
QY 632 NTPKEPAPTPEKAPPTPKAPFTTETAPTTPKGTAFTTLEKAPATTPKKPAKELA 691
Db 308 PSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 366
QY 692 PTTKEPTSTSDKAPATTGTAFTTPEKAPTTPEKAPTTGTAFTTLEKAPATTTP 751
Db 367 ---KSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 422
QY 752 KKPAPKELAPTTGPTSTSDKAPATTPEKLAFTTPEKAPTTPKKPAFTTPEETPP 809
Db 423 SPP-----PPYYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 475

```

Search completed: April 26, 2002, 16:32:49
 Job time: 764 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:32:55 ; Search time 62.75 seconds
(without alignments)
795.232 Million cell updates/sec

Title: AA7
Perfect score: 7323
Sequence: 1 MAMKTLPIYLILLSVFIQ.....ARATRRSGQLSKVWYNC 1361

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 segs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1194	16.3	5179	1 MUC2_HUMAN	Q02817 homo sapien
2	950	13.0	1664	1 SLPL_CLOTM	Q06852 clostridium
3	795.5	10.9	1367	1 ANYH_YEAST	P08640 saccharomyc
4	651	8.9	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
5	566.5	7.7	875	1 FPI_MYED	Q25460 mytilus edu
6	553.5	7.6	2700	1 ZAN_HUMAN	Q9Y493 homo sapien
7	531	7.5	620	1 EXTN_TOBAC	P13983 nicotiana t
8	533	7.3	1087	1 NFR_MOUSE	P19246 mus musculu
9	530.5	7.2	1162	1 TCNA_TRYCR	P23253 trypanosoma
10	530	7.2	865	1 CPN_DROME	Q02810 drosophila
11	518.5	7.1	872	1 FPI_MYTCCO	Q25434 mytilus cor
12	509.5	7.0	662	1 MUC1_XENLA	Q05049 xenopus lae
13	503.5	6.9	1970	1 RPB1_HUMAN	P24928 homo sapien
14	498.5	6.8	1970	1 RPB1_MOUSE	P08775 mus musculu
15	498	6.8	831	1 NFR_RAT	P16884 rattus norv
16	493.5	6.7	467	1 RPB1_CRIGR	P11414 cricetus
17	488.5	6.7	826	1 SSP2_PLAYO	Q01443 plasmodium
18	488.5	6.7	1020	1 NFR_HUMAN	P12036 homo sapien
19	475.5	6.5	267	1 EXTN_MAZE	P14918 zea mays (m
20	473.5	6.5	2142	1 BAT2_HUMAN	P48634 homo sapien
21	471.5	6.4	5376	1 ZAN_MOUSE	O88799 mus musculu
22	468.5	6.4	634	1 HWP1_CANAL	P46593 candida alb
23	467	6.4	817	1 VRP1_YEAST	P37370 saccharomyc
24	454	6.2	797	1 VGLX_HSVBR	P28968 equine herp
25	454	6.2	1161	1 Y3RP_YEAST	P47179 saccharomyc
26	448.5	6.1	670	1 VES0_HUMAN	Q00130 ictaluriid h
27	448.5	6.1	1229	1 N121_HUMAN	Q09243 homo sapien
28	442.5	6.0	751	1 FPI_MYTGA	O27409 mytilus gal
29	440	6.0	1794	1 FPI_SCHPO	O10172 schizosacch
30	439.5	6.0	1083	1 T2D3_HUMAN	O00266 homo sapien
31	438	6.0	2774	1 MAPA_RAT	P10920 rattus norv
32	434	5.9	3164	1 TEGU_HSV11	P14320 herpes simp
33	432.5	5.9	439	1 XP2_XENLA	P17437 xenopus lae

ALIGNMENTS

RESULT	ID	STANDARD	PRT	5179 AA
1	MUC2_HUMAN			
AC	002817	Q14878		
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).			
GN	MUC2 OR SMUC.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	[1]			
RC	SEQUENCE FROM N.A.			
RC	TISSUE=Intestine;			
RA	MEDLINE=94132002; PubMed=8300571;			
RA	Gum J.R., Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;			
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.			
RT	Identification of the amino terminus and overall sequence similarity			
RT	to prepro-von Willebrand factor.";			
RL	J. Biol. Chem. 269:2440-2446(1994).			
RN	[2]			
RP	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.			
RC	TISSUE=Colon;			
RA	MEDLINE=93016075; PubMed=1400449;			
RA	Gum J.R., Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,			
RT	Kim Y.S.;			
RT	"The human MUC2 intestinal mucin has cysteine-rich subdomains located			
RT	both upstream and downstream of its central repetitive region.";			
RL	J. Biol. Chem. 267:21375-21383(1992).			
RN	[3]			
RP	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.			
RC	MEDLINE=91358717; PubMed=1885763;			
RA	Toribara N.W., Gum J.R., Jr., Culhane P.J., Lagace R.E., Hicks J.W.,			
RT	Petersen G.M., Kim Y.S.;			
RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays			
RL	and polymorphism.";			
RL	J. Clin. Invest. 88:1005-1013(1991).			
CC	- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND			
CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A			
CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS			
CC	AGENTS AT MUCOSAL SURFACES.			
CC	- SUBUNIT: MULTIMERIC.			
CC	- SUBCELLULAR LOCATION: SECRETED.			
CC	- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,			
CC	BRONCHUS, CERVIX AND GALL BLADDER.			
CC	- PM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR			
CC	INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).			
CC	- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND			
CC	VARIABLE AMONG DIFFERENT ALLELES.			
CC	- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT			
CC	OF SILKWORM HEMOCYTIN.			
CC	- SIMILARITY: CONTAINS 2 WWFC DOMAINS.			
CC	- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).			

DR	EMBL; L21998; AAB955295.1;	-
DR	EMBL; M74027; AA559875.1;	-
DR	EMBL; M64131; AAA59163.1;	-
DR	EMBL; M94132; AAA59164.1;	-
DR	MIM; 158370;	-
DR	InterPro; IPR000359; Cys_knot.	.
DR	InterPro; IPR000561; EGF-like.	.
DR	InterPro; IPR002400; GE_Cysknot.	.
DR	InterPro; IPR001007; VMFC.	.
DR	InterPro; IPR001846; vwd.	.
DR	Pfam; PF000094; vwd; 4.	.
DR	Pfam; PF000943; GFCYSKNOT.	.
DR	PRINTS; SMO0214; VMC; 2.	.
DR	SMART; SMO0011; VMC_def; 2.	.
DR	SMART; SMO0216; VWD; 4.	.
DR	PROSITE; PS00022; EGF_1; UNKNOWN_1.	.
DR	PROSITE; PS01185; CTCK_1; 1.	.
DR	PROSITE; PS01225; CTCK_2; 1.	.
DR	PROSITE; PS01208; VMFC; 2.	.
KW	Glycoprotein; Repeat; signal.	.
FT	SIGNAL	1 20 POTENTIAL.
FT	CHAIN	21 5179 MOCN 2.
FT	DOMAIN	1401 1747 APPROXIMATE REPEATS.
FT	REPEAT	1401 1416 1.
FT	REPEAT	1417 1432 2.
FT	REPEAT	1433 1448 3.
FT	REPEAT	1449 1464 4.
FT	REPEAT	1465 1471 5.
FT	REPEAT	1472 1478 6.
FT	REPEAT	1479 1494 7A.
FT	REPEAT	1495 1517 7B.
FT	REPEAT	1518 1533 8A.
FT	REPEAT	1534 1556 8B.
FT	REPEAT	1557 1572 9A.
FT	REPEAT	1573 1596 9B.
FT	REPEAT	1597 1612 10A.
FT	REPEAT	1613 1635 10B.
FT	REPEAT	1636 1651 11A.
FT	REPEAT	1652 1675 11B.
FT	REPEAT	1676 1683 12.
FT	REPEAT	1684 1699 13.
FT	REPEAT	1700 1715 14.
FT	REPEAT	1716 1731 15.
FT	REPEAT	1732 1747 16.
FT	DOMAIN	4815 4886 VMFC 1.
FT	DOMAIN	4924 4991 VMFC 2.
FT	DOMAIN	5075 5160 CTCK.
FT	DISULFID	5075 5122 BY SIMILARITY.
FT	DISULFID	5089 5136 BY SIMILARITY.
FT	DISULFID	5098 5152 BY SIMILARITY.
FT	DISULFID	5102 5154 BY SIMILARITY.
FT	DISULFID	? 5159 BY SIMILARITY.
FT	CARBOHYD	163 163 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	670 670 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	770 770 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	894 894 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1139 1139 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1134 1134 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1215 1215 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1230 1230 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1246 1246 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1787 1787 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1820 1820 N-LINKED (GLCNAC. .) (POTENTIAL).

FT	CARBOHYD	4339	4339	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4351	4351	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4362	4362	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4373	4373	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4422	4422	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4438	4438	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4502	4502	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4616	4616	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4627	4627	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4752	4752	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4787	4787	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4881	4881	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4888	4888	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4955	4955	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4970	4970	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	5019	5019	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	5038	5038	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	5069	5069	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CONFLICT	1351	1351	H -> L (IN REF. 3).		
FT	CONFLICT	1412	1412	T -> S (IN REF. 3).		
FT	CONFLICT	1449	1449	L -> P (IN REF. 3).		
FT	CONFLICT	1504	1504	M -> T (IN REF. 3).		
FT	CONFLICT	4192	4192	G -> S (IN REF. 2).		
SQ	SEQUENCE	5179 AA;	540295 MW;	85CD7571FB9A5663 CRC64;		
<hr/>						
Query Match		16.3%;	Score 1194;	DB 1;	Length 5179;	
Best Local Similarity		30.6%;	Pred. No. 5,-6e-40;			
Matches 400;		Conservative	74;	Mismatches 515;	Indels 320;	Gaps 51.
OY	4	KLLPYLILLLEVFYIQOVSODLSGC--AGKCG-----EGSRBQATCDXNCOHUYMCC	56			
Db	1169	KRPRII-----EEDLKCVYADCKGVEDTHTRPGASVTEETCKSCVC-C	1213			
OY	57	CPDFKVCYAEILSCKRCFSEFERG----RBCDDACKKYDKCC---PDYESFCAL	106			
Db	1214	TNSGYVNCPE---BEKILNQTDGAFCYMWELCGRGNGVEKHFNIDISITRRSTLTFTT	1270			
OY	107	VHNPSPPS-SKKAPPGCASQTIKSTTKRSRPRPKKKTKKVISEELITVKDNKKNT	165			
Db	1271	ILPLPTTFRTTTTTTTTTTPSTSVISTT--PK-----LCCLMSDWINE-----	1311			
OY	166	KKKPTRKPRVDEAGSGSLNGDFK-----VTTPDRNSTQH-NKVSTSPK	208			
Db	1312	-----DHRSSGDSDGDERPDGVCGAFEDIKCSVKRPHLSLBQHCQKVCDSV	1360			
OY	209	-----TTAKPINRPSLRPNSTSKETSL	233			
Db	1361	VGFICKNEODFGNGPFLCYDYKIRVNCCMPMDKCJITTPSPPTTSSPPTTTTLDPPT	1420			
OY	234	TVNKETTVKEKENTTNKOSTDCKEKTSASKETQSILEKTSADKLAPYSKVLAKPTPKAE	293			
Db	1421	TPSPPTT-----TTTTPPPTTTPSPPTTTTTTP-----LPIT-----TPSPDIS	1459			
OY	294	TTTGCFALTTPKREPT-----PTTPKEBASTTPKEPPTTIKSAP-TTPKEP-----AP	340			
Db	1460	TTTTPTPTTTPSPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPSPPMYTPIIPPASTTLPP	1519			
OY	341	TTTKSAP-TTPKEAPPTTKEAPTPKPE-----APTTPKEAPPTTKSAP--TTPK	390			
Db	1520	TTTTSPPTTTTTTPPTTTPSPPTTTPPTTTPPTTTLPLPTTTPSPPTTTPPTTTPPS	1579			
OY	391	EPAETPKRKAPPTPKEPAPTTPKEPPTTPPKEPAPPTTKEPAPPTPKPEP-----APTA	443			
Db	1580	PPPTTTPSPPTTTTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT	1639			
OY	444	PKRPAPTTPKEAPPTPKEPAPTTPKEBSPTTPKEAPPTTKSAPTTTKCAPPTTKSAP	503			
Db	1640	TPSPPTTTPPTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPSPPTTTPPTTTP	1699			
OY	504	---TTKEBPPTTTPKEAPPTPKEPAPTTPKKAPAPTTPPKAPAPTTPTKKKAAPT	560			
Db	1700	SPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT	1758			

```

QY 561 AKREPAATP-----KETATP-----578
| : | | | |
DB 1759 TPLPSTPTTFSPTSTTTTTCVPLCNMTGNLDSKPNFKRGDTLIGVCGRGMA 1818
QY 579 -----KLLP-----583
| : |
DB 1819 ANISCRATMPDVPICGLQGVWCDVSVGLICKNEDKPGVGIIPMAFCLNVEINVCCEC 1878
QY 584 -TPREKAPATTPKPAATPTEBELATPTEPTPTP-TPREAPATTPKAAATKREAPATP 641
| : | | | | | | | | | |
DB 1879 VTQPTTMTTTEBPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTG 1938
QY 642 PKBP-----APTPKREAPAT-TPKETAATTPKGTATPTTLKEPATTPKPKAPKELATPTT 695
| : | | | | | | | | | |
DB 1939 TQPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 1998
QY 696 KEPTSTTSKAPAP-TPPKGTATPTTPKREAPATTPKGTATPTTLKEPATTPPKP 754
| : | | | | | | | | | |
DB 1999 VTPTPTPTGQPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTGQTP 2057
QY 755 APKELATPTTKGPTSTTSKAPAP-TPPKETAATTPKREAPATTPKPKAPATPTEPTPTSE 813
| : | | | | | | | | | |
DB 2058 TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2116
QY 814 VSPPTTKPEPTTIHKSPEDESTPELSAEPKALENSEKPEP-----GVPT-TPKTPA 862
| : | : | : | : | : | : | : | : |
DB 2117 TPTPTGQPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2173
QY 863 ATPEKMTTKAKDKTTERDLR---TTP-ETTPAAPKWT---KETATTEKTESKITATP 914
| : | | | | | | | | | |
DB 2174 TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2233
QY 915 TQVSTTTTQDTPEPKITTLTKTLTAPKVTTT-KLITTEIMKPEETAKPKRATNRSKA 973
| : | | | | | | | | | |
DB 2234 TP-TGQTPPTTP--ITT--TTTVPPTPTPTGQTPPTPTPTPTPTPTPTPTPTPTPTPTPT 2288
QY 974 TTPPKPKATPAP-KKPTSTKKKPTMPVRKKTPTPKMTSTMPBLNP---TSRIEAM 1029
| : | | | | | | | | | |
DB 2289 TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2348
QY 1030 LOTTTR-PNQTSPKSLVENVPKSEDAGAGCEPHMLRNHVMPETPT 1077
| : | : | : | : | : | : |
DB 2349 TPTPTGQPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2392

```

```

RESULT 2
SLIP1_CLOTH STANDARD: PRT: 1664 AA.
AC 006852:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CELL SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER
  PROTEIN 1).
GN OLPB.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 10682;
RA MEDLINE=9320931; PubMed=8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
  cellulosomal scaffolding protein ClpA and a protein possibly involved
  in attachment of the cellulosome to the cell surface."
RT J. Bacteriol. 175:1891-1899(1993).
CC -1- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.
CC -1- SUBCELLULAR LOCATION: CELL WALL.
CC -1- SIMILARITY: CONTAINS 4 S-LAYER HOMOLOGY (SLH) DOMAINS.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation-CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

DR EMBL: X67506; CAA47841.1; -.
DR InterPro: IPR001119; SLH.
DR Pfam: PF00395; SLH; 3.
DR PROSITE: PS01072; SLH DOMAIN; 2.
KW Cell wall; S-layer; Signal; Repeat.
FT SIGNAL 28
FT CHAIN 1
FT DOMAIN 29 1664 CELL SURFACE GLYCOPROTEIN 1.
FT REPEAT 36 763 4 x 156 AA APPROXIMATE REPEATS.
FT REPEAT 36 191 1.
FT REPEAT 207 363 2.
FT REPEAT 409 565 3.
FT REPEAT 607 763 4.
FT DOMAIN 771 1377 APPROXIMATE TANDEM REPEATS OF
  T-P-S-D-E-P.
FT DOMAIN 1378 1449 GLY/PRO/SER/THR-RICH.
FT DOMAIN 1453 1494 SLH 1 (INCOMPLETE).
FT DOMAIN 1495 1565 SLH 2.
FT DOMAIN 1566 1625 SLH 3.
FT DOMAIN 1626 1646 SLH 4 (INCOMPLETE).
SQ SEQUENCE 1664 AA; 178194 MW; 5F396695BA9FE74B CRC64;

```

Query Match 13.0%; Score 950; DB 1; Length 1664;
 Best Local Similarity 31.0%; Pred. No. 7,9e-31;
 Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;

```

QY 284 VLAKPTP-KAETTKGPAATTPKEPTPTTPKREAPATTPKKEPTPTTKSAPTPKREAPATP 342
| : | : | : | : | : | : | : | : |
DB 758 VVIQPAPIKAASDEPIPTDPSDEPTPS-----DEPTPS---DEPTPSDEPTPSD 804
QY 343 TKSAPTPKREAPATTPKREAPATTPKREPA-----PTTKKPAATTKSAPTPKREAPATP 398
| : | : | : | : | : | : | : | : |
DB 805 EPTPSETPEEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 864
QY 399 KPAATTPKREAPATTPKREAPATTPKREAPATTPKREAPATTPKREAPATTPKREAPATP 457
| : | : | : | : | : | : | : | : |
DB 865 SDEPTPSDEPTPS--DEPTPS--DEPTPS--DEPTPS--DEPTPS--DEPTPS--DEPTPS 919
QY 458 TTPKREPA--PTTKKESPT-TPKREAPATTKS-APTTPKREAPAT--TTKSAPTPKES 510
| : | : | : | : | : | : | : | : |
DB 920 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 979
QY 511 ----PTTKKREAPAT-TPKREAPATTPKPKAPATTPKREAPATTPKREAPATTPKREAPATP 565
| : | : | : | : | : | : | : | : |
DB 980 PSEPTPSDEPTPSSEPTPEEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1039
QY 566 APPT---TPKETAATTPKPKLPTTPKELAPATTPKREAPAT--TPBELAPATPEEPTPT-TP 618
| : | : | : | : | : | : | : | : |
DB 1040 TPEDEPTPSE---TPBELIPTDPSDEPTPSDEPTPSDEPTPSD-EPTPSDEPTPSSEPT 1094
QY 619 EEPAPTPPKAANPTKREAPATTPKREPA--PTTKKREAPAT-TPKETAATTPKGTATPTTK 675
| : | : | : | : | : | : | : | : |
DB 1095 EEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD 1154
QY 676 EEPATTPPKKAPKELAPATTPKKEPT-STTSKAPATTPKGTATTPKREAPATTPKREAPATP 734
| : | : | : | : | : | : | : | : |
DB 1155 EPTPS--DEPTPSD-EPTPSDEPTPSSEPTPEEPIPTDPSDEPTPSDEPTPS--DEPTPS- 1208
QY 735 PKGTATTPKREAPAT-TPPKKAPKELAPATTPKSTTSKAPATP--TPKETAATTPKRE 790
| : | : | : | : | : | : | : | : |
DB 1209 ---DEPTPSDEPTPSSEPTPEEPI-----EPTDPSDEPTPSDEPTPSD-EPTPSDE 1253
QY 791 PAPTTPKKAPATP-TPETPTPTSEVSTPTTKKEPTTIHKSPEDESTPELSAEPKALENS 849
| : | : | : | : | : | : | : | : |
DB 1254 PTPS--DEPTPSSEPTPEEPIPTDPSDEPTPSDEPTPS---PSDEPTP--SDEPTPSDEPTP 1305

```



```

Db 694 PVT-----SSTESSAAVPPSSSTSSSAVPPPSSSTESSAAVPPPS-ASTESS 748
Qy 706 PAPTPKGAATTPKKEPAP---TPKKEPAPTPPKGAATTPKKEPAPTPPKAPKELAP 761
Db 749 SAPVTSSSTESSAAVPPPPSSSTESSAAVPPPPSSSTESSAAVPPPPSSSTESSAAV 808
Qy 762 TTTGCPSTSDKAPAPTPKKEPAP---TPPKAPAPTPPPPTTSEVST 816
Db 809 VPPSSSSNTSSAPSSSTESSAAVPPPPSSSTESSAAVPPSSSTESSAAVPP 868
Qy 817 PT-----TKKPTTHKSPDESPELSAEPKALENSPKKPGVPTTKTPAATPEMTT 871
Db 869 PSSSSNTSSAPSSI---PPSSSTESSSTGT-TVPSSSKIPGSGOTETSVSSTETIIV 924
Qy 872 AKDKTERDLRTPEPTTAAAPKMTKETATTEKTESKITAT---TQVSTTTQDTTIPK 929
Db 925 TKTTSTVTPSTTTTITTVOSTGNSAGETSSGSPKTVTTTITTTTSTTSTTTIT 984
Qy 930 ITTKKTTTLAPKVT---TKKTTT---EIMNKPPEYAKPKDRATNSKATTPKPKP- 981
Db 985 TTVOSTGNSAGETSSGSPKTTTITTVPCSTPSEEA-----SESTTSTPTPTTV 1036
Qy 982 -----TKAKKPTSTKKPKTTPRVRKPKTTPPKMTSMPELNPSTRIAE 1028
Db 1037 VSTTVVTEYESTKPGCEITTFVTKNIPPTTITTAIPR--SVTVVTFPTPT-ITTP 1093
Qy 1029 MLOTTPRNOTPNSKLEVPKSEDAAGETPPHMLRHPVPEVTPMDYLPBPVNO 1088
Db 1094 VCSGT-----NSAGETSGCSPKVTITTVPCSTGGETTEATLVTAAVT 1140
Qy 1089 GIINPLMSEETINCKGPDVGLTTLNGLVAFRGHYEMMLSEPPSPARRITEVGI 1148
Db 1141 TTVVTESSSTGNSA-GKITTYGTVKSVPTT-----YVTLAASAVTBPATN-----AV 1188
Qy 1149 PSPIDVETRC---NCEGKT 1165
Db 1189 PTTITT--TECSAATNAGET 1207

RESULT 4
MUC1_HUMAN STANDARD: PRT: 1255 AA.
ID MUC1_HUMAN P15942, P13931, P17626, Q14128, Q16442, Q16437, Q9Y4J2;
AC P15941, P15942, P13931, P17626, Q14128, Q16442, Q16437, Q9Y4J2;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM) (PEMT)
DE (EPITHELIAL) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)
DE (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H23A6) (PEANUT-
DE REACTIVE URINARY MUCIN) (PUM) (BREAST CARCINOMA-ASSOCIATED ANTIGEN
DE DF3).
GN MUC1.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE-Pancreas;
RX MEDLINE=90368716; PubMed=2394722;
RA Ian M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL J. Biol. Chem. 265:15294-15299(1990).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=90202794; PubMed=2318825;
RA Lichtenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT "Epistatin, a carcinoma-associated mucin, is generated by a
RT polymorphic gene encoding splice variants with alternative amino
RT termini.";
RL J. Biol. Chem. 265:5573-5578(1990).
RN [3]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE-Breast carcinoma;
RX MEDLINE=90368715; PubMed=1697589;
RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA Bat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of human tumor-associated
RT polymorphic epithelial mucin.";
RL J. Biol. Chem. 265:15286-15293(1990).
RN [4]
RN SEQUENCE FROM N.A.
RX MEDLINE=91097524; PubMed=2268309;
RA Lancaster C.A., Bat N., Duhig T., Wilson D.,
RA Taylor-Papadimitriou J., Gendler S.J.;
RT "Structure and expression of the human polymorphic epithelial mucin
RT gene: an expressed VNIR unit.";
RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN [5]
RN SEQUENCE FROM N.A.
RC TISSUE-Breast carcinoma;
RX MEDLINE=90276413; PubMed=2351132;
RA Wreschener D.H., Hareven M., Tsarfaty I., Smorodinsky N.,
RA Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,
RA Keydar I.;
RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms.";
RL Eur. J. Biochem. 189:463-473(1990).
RN [6]
RN SEQUENCE FROM N.A.
RC TISSUE-Breast carcinoma;
RX MEDLINE=90276414; PubMed=2112460;
RA Hareven M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
RA Zrihen S., Weiss M., Green S., Lathe R., Keydar I., Wreschener D.H.;
RT "A transcribed gene, containing a variable number of tandem repeats,
RT codes for a human epithelial tumor antigen. cDNA cloning, expression
RT of the transfected gene and over expression in breast cancer
RT tissue.";
RL Eur. J. Biochem. 189:475-486(1990).
RN [7]
RN SEQUENCE FROM N.A.
RX MEDLINE=91033045; PubMed=1688329;
RA Tsarfaty I., Hareven M., Horev J., Zaretsky J., Weiss M.,
RA Jeltsch J.M., Gartner J.M., Lathe R., Keydar I., Wreschener D.H.;
RT "Isolation and characterization of an expressed hypervariable gene
RT coding for a breast-cancer-associated antigen.";
RL Gene 93:313-318(1990).
RN [8]
RN PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=8830762; PubMed=3417635;
RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA Burchell J.;
RT "A highly immunogenic region of a human polymorphic epithelial mucin
RT expressed by carcinomas is made up of tandem repeats.";
RL J. Biol. Chem. 263:12820-12823(1988).
RN [9]
RN SEQUENCE OF 1-169 FROM N.A.
RX MEDLINE=90088473; PubMed=2597151;
RA Abe M., Sidiqul J., Kufo D.;
RT "Sequence analysis of the 5' region of the human DF3 breast
RT carcinoma-associated antigen gene.";
RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN [10]
RN SEQUENCE OF 1-109 FROM N.A.
RC TISSUE-Thyroid;
RX MEDLINE=96183746; PubMed=8608966;
RA Weiss M., Baruch A., Keydar I., Wreschener D.H.;
RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT transcriptase polymerase chain reaction of the MUC1 gene.";
RL Int. J. Cancer 66:55-59(1996).
RN [11]
RN SEQUENCE OF 1-89 FROM N.A.
RC TISSUE-Lung;
RX MEDLINE=96181716; PubMed=8604237;
RA Yu C.-J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,

```

RA Lee L.N., Luh K.T., Wu C.W.;
 RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
 tissues.";
 RN Oncology 53:118-126(1996).
 RC SEQUENCE OF 1-46 FROM N.A.
 RA TISSUE-Breast carcinoma;
 RA Buluwela L., Liu Q., Lugmani Y.A., Gomm J.J., Coombes R.C.;
 RL Submitted (Oct-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
 CYTOSKELETON.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM
 IS ALSO PRODUCED.
 CC -1- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE
 SPLICING.
 CC -1- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL
 TUMORS, SUCH AS BREAST CANCER.
 CC -1- PPM: HIGHLY GLYCOSYLATED (N-AND O-LINKED CARBOHYDRATES AND SIALIC
 ACID).
 CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT
 VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE
 MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.
 CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J05582; AAA60019.1; -
 DR EMBL: M32738; AAA35804.1; -
 DR EMBL: M32739; AAA35806.1; -
 DR EMBL: J05581; AAA59876.1; -
 DR EMBL: M61170; AAB53150.1; -
 DR EMBL: X52229; CA364478.1; ALT_SEQ.
 DR EMBL: X52228; CA364477.1; ALT_SEQ.
 DR EMBL: M35093; AAB59612.1; ALT_SEQ.
 DR EMBL: Z17324; CA78972.1; -
 DR EMBL: Z17325; CA78973.1; -
 DR EMBL: M31823; AAA35757.1; -
 DR EMBL: S81781; AAD14376.1; ALT_INT.
 DR EMBL: S81736; AAD14369.1; ALT_INT.
 DR EMBL: M21868; AAA59874.1; ALT_SEQ.
 DR PIR: A35175; A35175.
 DR PIR: B35175; B35175.
 DR PIR: S10218; S10218.
 DR GLYCOSULEDB: P15941; -
 DR MIM: 158340; -
 DR MIM: 113720; -
 DR InterPro: IPR000082; SEA.
 DR PIRam: PF01390; SEA; 1.
 DR SMART: SMO0200; SEA; 1.
 DR PROSITE: PSS0024; SEA; 1.
 KW Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
 KW Repeat: Alternative splicing.
 FT SIGNAL 1 23
 FT CHAIN 24 1255
 FT DOMAIN 24 1162
 FT TRANSMEM 1163 1255
 FT DOMAIN 81 960
 FT DOMAIN 1034 1151
 FT CARBOHYD 957 957
 FT CARBOHYD 975 975
 FT CARBOHYD 1029 1029
 FT CARBOHYD 1055 1055
 FT CARBOHYD 1133 1133
 FT VARSPPLIC 19 19
 FT VARSPPLIC 20 22
 FT VARSPPLIC 20 31
 FT VARSPPLIC MISSING (IN ISOFORM C).

FT VARSPPLIC 126 905 MISSING (IN POLYMORPHIC EPITHELIAL
 FT VARSPPLIC 1077 1087 ISOFORM).
 FT VARSPPLIC 1088 1255 FLUOROKOGL -> VSLGSLPMLP (IN SECRETED
 FT VARSPPLIC 1088 1255 ISOFORM).
 FT VARSPPLIC 1088 1255 MISSING (IN SECRETED ISOFORM).
 FT CONFLICT 2 2 T -> A (IN REF. 11).
 FT CONFLICT 134 134 P -> Q (IN REF. 9).
 FT CONFLICT 154 154 P -> Q (IN REF. 9).
 FT CONFLICT 1021 1021 S -> T (IN REF. 3).
 FT CONFLICT 1251 1251 A -> T (IN REF. 3).
 SQ SEQUENCE 1255 AA; 122072 MW; 5E2BDC4DEF7D9A82 CRC64;
 Query Match 8.9%; Score 651; DB 1; Length 1255;
 Best Local Similarity 27.8%; Pred. No. 3,2e-19;
 Matches 295; Conservative 72; Mismatches 460; Indels 234; Gaps 51;
 231 TSLTVNKKETVEMKETTNNKOTSDGKEKTSKAKTOSIEKTSKAD-LAPTSKVLAKPT 289
 16 TLVLV-----VTSGHASSTPGCEKETATORTSSVSTREKNAVMTSSLSHS 65
 290 P-KAETTKGP--ALTTPKEP-----TPTPKEPASTTPKKEPTPTT---K 329
 66 PGSGSSTTGQDVTAPATPEPAGSAAATWGQDVTVPTRPALGSTPPAHADVTSAPDNK 125
 330 SAPTPKEPAPTTKSAPTPKKEPAPTTKPEP-----APTPKEPAPTTKKEPAPTTK 383
 126 PAPSSTAPPAHGVTSAPDT--RAPSTAPPAHGVTSAPDT--RAPSTAPPAHGVTS 179
 384 SAPTPKEPAPTTK-----KKRAPTPKEP-----APTPKEPPTPTT----- 421
 180 SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGV 239
 422 ----KEPAPTTKEP-----APTPKEPAPTTAK-----KRAPTPKEPAPTTPK-----E 462
 240 SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGV 299
 463 PAPTTPKEPSTTPKKEPAPTT--TKSAPTTKKEPAPTTKSAPTPKESPTTKKEP-- 517
 300 SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAH 356
 518 ----APTPKEPAPTTK-----KRAPTPKEPAPTTK-----EPAPTTKKEPAPTTAKEP 565
 357 GVTAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAH 416
 566 APTPKKEPAPTTKPKLPTTPPEKAPTTPEKAPTTPEELAPTTPEPPTTPPEAPPT 625
 417 GVTAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAP--PAHGVTSAPDTRPAPGSTA 472
 626 PKA----AAPTPKEPAPTTK-----EPAPTTKKEPAPTTKETAPTPKGTAPTTKKEP 677
 473 PPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAP--PAHGVTSAPDTRPAP 528
 678 APTP-----KKRAPKELAPTTKKEPTSTSKRAP--TPPKGTAPTTKKEPAPT 725
 529 GSTAPPAHGVTSAPDTRPAPGSTAP--PAHGVTSAPDTRPAPGSTAPPAHGVTS---AP 583
 726 TPKEPAPTTKPG-----TAPTTKKEPAPTT-----KKRAPKELAPTTKPGTST 771
 584 TRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAP--PAHGVTSAP 642
 772 SDKRAP--TPPK-----ETAPTTKKEPAPTT-----KKRAPPTPETPTPTSEVS 815
 643 DTRAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGST--TAPPAHGVTS 700
 816 TPTTTKEPT-----THKSPDESTPELSAPTPKALENSPKKEPVTTTPATKPEM 868
 701 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAP--DTRPAPGSTAP-- 753
 869 TTPKDKTTEEDLRTPTPTT-----TAAPKMKETATTEKTEKTKITATTOVSTT 921
 754 --PAHGVTSAPDTRPAPGSTAPPAHGVTSAP--DTRPAPGSTAPPAHGVTSAPDTRPA 807

RA Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,
 RA Prioli R.P.;
 RT "The Trypanosoma cruzi neuraminidase contains sequences similar to
 RT bacterial neuraminidases, WTD repeats of the low density lipoprotein
 RT receptor, and type III modules of fibronectin."
 RL J. Exp. Med. 174:179-191(1991).
 RN [2]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE-91376547; PubMed-1896773;
 RA Prioli R.P., Mejia J.S., Ajl T., Akawa M., Pereira M.E.A.;
 RT "Trypanosoma cruzi: localization of neuraminidase on the surface of
 RT trypomastigotes."
 RL Trop. Med. Parasitol. 42:146-150(1991).
 CC -1- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN
 CC PARASITE INVASION OF CELLS.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC
 CC LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACETYLNEURAMINYL
 CC RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACETLATED
 CC NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,
 CC GLYCOLIPIDS OR COLOMINIC ACID.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POSSIBLE).
 CC -1- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPOMASTIGOTES, MINIMUM
 CC IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.
 CC -1- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT
 CC DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TNCA
 CC PROTEIN.
 CC -1- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.
 CC -1- SIMILARITY: CONTAINS 3 BNR REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL: M61732; AAA30255.1; -
 CC DR PIR: JH0557; JH0557.
 CC DR HSP: P29768; IDIU.
 CC DR InterPro: IPR002860; BNR.
 CC DR Pfam: PF02012; BNR; 2.
 CC KM Hydrolase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;
 CC KW Phosphorylation.
 CC FT DOMAIN 1 457 CYS-RICH.
 CC FT REPEAT 23 34 BNR 1.
 CC FT REPEAT 163 174 BNR 2.
 CC FT REPEAT 209 220 BNR 3.
 CC FT DOMAIN 458 588 FIBRONECTIN TYPE-III.
 CC FT DOMAIN 589 1120 44 X 12 AA TANDEM REPEATS. LTR DOMAIN.
 CC FT CARBOHYD 342 342 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SQ SEQUENCE 1162 AA; 120032 MW; 07049221897C6A40 CRC64;

Query Match 7.2%; Score 530.5; DB 1; Length 1162;
 Best Local Similarity 29.4%; Pred. No. 1.6e-14;
 Matches 177; Conservative 100; Mismatches 212; Indels 113; Gaps 34;

QY 308 TPTTPEKE-PASTTPEKEPTTTTSAPTTPEKEPAATTTKSAPTTPEKEPAATTTTKEPAATTP 366
 DB 599 TPTTPEKEPAATTTTSAPTTPEKE-PAATTTPEKEPAATTTTKEPAATTTTKEPAATTP 413
 QY 367 KE-PAATTTTKEPAATTTTSAPTTPEKE-PAATTTPEKEPAATTTTKEPAATTTTKEPAATTP 413
 DB 652 ADSAAGTGTSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPY 711
 QY 414 KEPTT-PTTKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTP 472
 DB 712 VDSAGTGTSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPY 764

QY 473 PTTTKEPAATTTTSAPTTPEKEPAATTTTSAPTTPEKEPSPTTKEPAATTTPEKEPA-PT 528
 DB 765 HGTSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHS 817
 QY 529 TP-KKPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKEPA 563
 DB 818 TPVDSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPY 876
 QY 584 TPTTPEKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKE 633
 DB 877 TPVDSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPY 935
 QY 634 PKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTP 693
 DB 936 PSTPADSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSST 963
 QY 694 TTEPTTSTSDKPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTP 752
 DB 984 -PSTPADSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHS 1033
 QY 753 KPAPKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKEPAATTTTKEPA 808
 DB 1034 TP-----ADSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPY 1084
 QY 809 PTTSEVSTPTTKEPTTTTKEPTTTTKEPTTTTKEPTTTTKEPTTTTKEPTTTTKEPTTTTKEPT 867
 DB 1085 DSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPYDSSAHSSTPY 1125
 QY 868 WT 869
 DB 1126 GT 1127

RESULT 10
 ID CPN_DROME STANDARD; PRT; 865 AA.
 AC 002910;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE CALPHOTIN.
 GN CPN OR CAP.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-CANTON-S;
 RX MEDLINE-93165730; PubMed-8434015;
 RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
 RT calcium and contains a leucine zipper."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
 CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
 CC REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
 CC OF CA+2 PER MOL OF PROTEIN.
 CC -1- SUBUNIT: HOMODIMER (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
 CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
 CC COMPOUND EYES AND OCELLI.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
 CC DEVELOPMENT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: L02111; AAA28405.1; -
 DR EMBL: L05080; AAA28420.1; -
 DR PIR: A47282; A47282.
 DR FlyBase: FBgn0010218; Cpn.

KW Calcium-binding.
 FT CONFLICT 36 A -> AVAPAVVA (IN REF. 2).
 FT CONFLICT 43 I -> T (IN REF. 2).
 FT CONFLICT 64 I -> V (IN REF. 2).
 FT CONFLICT 76 T -> A (IN REF. 2).
 FT CONFLICT 100 P -> PP (IN REF. 2).
 FT CONFLICT 126 VQ -> AP (IN REF. 2).
 FT CONFLICT 154 I -> V (IN REF. 2).
 FT CONFLICT 160 S -> T (IN REF. 2).
 FT CONFLICT 534 A -> E (IN REF. 2).
 FT CONFLICT 699 I -> T (IN REF. 2).
 FT CONFLICT 703 V -> L (IN REF. 2).
 FT CONFLICT 721 D -> E (IN REF. 2).
 SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFC CRC64;

Query Match 7.28; Score 530; DB 1; Length 865;

Best Local Similarity 26.28; Pred. No. 1.3e-14;
 Matches 233; Conservative 73; Mismatches 433; Indels 152; Gaps 41;

OY 279 APTSKVLAKP-TPK---AETTKCPALTPPKPEPTTPPKPEASTTPPKPEPTTKSAPTT 334
 DB 8 SPVSAPVAAVTBSAVALPQVVSAPVAP-----APAPVATPPVAPPTTASVCPAT 61
 OY 335 PKPEAPPTTKSAPTTP-KEBAPPTTKPEAPPTTKPEAPPTTKKEBAPT---TTKSADPTT 389
 DB 62 VTIPAPAPVIAASVTVASVAPVVAAPTP-----PAASPTSTPVAAQIIPVAVSAPVAP 116
 OY 390 KEAPAPTPKKRPAP-----TTKPEAPPTTKPEPTP-----TTKPEAPPTTKPEAPPT 435
 DB 117 PVATPTPVVQIIPVAPVIAVTPVPAASAPTPVAPVAPVIAVSPVAPVAPVAPVAPV 176
 OY 436 PKPEAP-PTAPKKRPAPPTTKPEAPPTTKPEAPPTTKPE-PSPTTPKEBAP-----TTKSA 487
 DB 177 AAVPAAPVAVVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVATK 236
 OY 488 PTTTKPE---APTTKSAPTTKPEPTTKPEAP-PTTKPEAPPTTKKKRPAPPTTKPEAP 543
 DB 237 PLAAAEVVAAPVAPTEPVVAPPAASPHVAVAPVAVAPVAVASAPTEPPVAAATLTTPAP 296
 OY 544 TTP-----KEBAPPTTKKPAPTKAPEP---APTTPKET-----APTTPKKLTPT 584
 DB 297 ETPLAPVAVASOVAAVNTVATPTTPAPPEPTTAPVPAVPAETPEVASVAVAPETTPPVVPV 356
 OY 585 TPPEKL-APTTPKEKAPPTTPEELAPTTPEEPTPTTPEEP---APTTPKAAAPNTPKPEAPT 640
 DB 357 AASISAPVAVATTPVAPAT---LAVTDDVTAASVAPBELPVVAPSPVSAVAETPVDLAP 413
 OY 641 TPKEAPPTTKPEAPPTTKETAPTPP-KGTAPTTLKE-----PAPTTPKKAPKE 689
 DB 414 V---LPVAAEPVPAVAAETPETPAPASAPVIAALDPEVAPVIAASDAPAEAPSA 469
 OY 690 LAPTTKEPTSTSDKAPPTPKGTAPTP-----KEBAPPTPKPEP---APTTPKGTAPT 741
 DB 470 AAPVAVSTPP---TTSVETTAAPRAVAPTEPIDVSVLSEAAETPVAPVEVETTVANADV 527
 OY 742 TLKEPA-----PTTPKKAPKELAPTTT---KGTSTSTSDKRAPTT---PKETAPTT 787
 DB 528 APPEAAADLIIEPEPAPVAPIDLEQTTSVPAVEAAESTSS--PIPETSLPPPEAAVAP 585
 OY 788 PKPEAPPTTKKPAPPTTETPTPTTSEVSTPTTKEPTTKSPD-----ESTPEL 837

DB 586 EVAAVAPITAPPEIP-----EPEPSLATPTPEIPVAPVAPVIOAVDAVEVPTETSTIPET 641
 OY 838 SAETPKALENSPKPEGVPTTKTPATKPKPEPTTKAKTKTERDRTTP--ETTAAAKMT 895
 DB 642 TVE-FPRAVAKVVDPAI--TEAPVTOEDPDVANINGAPAPATEL-TTPVAEIVTAAAEVS 697
 OY 896 KEVATTEKTESKITPTTTTOVST-----TTODTTPFKITTLTKTTLA-PKVTTTKKT 948
 DB 698 DIAIPVDDPPVPOEINAAIEPDTKPAEIVVESTIPIEAPVPEVSKVAPVISEAPAA 757
 OY 949 ITTEINMKPEETAKPKDRATNSKATTPPKPKPAKPKPTSTPKPTMPRVKPKTPTT 1008
 DB 758 EVPTTADNDPNTSVGISSEVPTIAEKPVVEPPTS--EIEQSSSPS--DSVPVAKITPL 813
 OY 1009 PKRKTSTMPELNPTSRTAEAM---LQTTTRNQNPNKSLVAVN-----PKS 1051
 DB 814 LRDIQTTDVSLAIAATLDAIGERLKDQKARNOQVMDRLCEIEKILGPPKS 864

RESULT 11
 FPL_MYTCO
 ID FPL_MYTCO STANDARD; PRT; 872 AA.
 AC Q25434;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR (FOOT PROTEIN 1) (MCPPI).
 GN FPL.
 OS Mytilus coruscus (Sea mussel).
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
 OC Mytiloidae; Mytilidae; Mytilus.
 OC NCBI_Taxid=42192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FOOT.
 RX MDLLINE=96394686; PubMed=8798340;
 RA Inoue K., Takeuchi Y., Takeyama S., Yamaha E., Yamazaki F., Odo S.,
 RA Harayama S.;
 RT "Adhesive protein cDNA sequence of the mussel Mytilus coruscus and
 RT its evolutionary implications.";
 RL J. Mol. Evol. 43:348-356(1996).
 CC - FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
 CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
 CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
 CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
 CC - SUBCELLULAR LOCATION: SECRETED.
 CC - TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
 CC - DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
 CC - PTM: PROLINES IN THE REPEAT ARE HYDROXYLATED (SINGLE OR DI-) AND
 CC ALSO TYROSINE (THUS PRODUCING DOPA = 3,4-DIHYDROXYPHENYLALANINE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: D63777; BAA09850.1; -
 DR InterPro: IPR002964; Adhesive_plaq.
 DR InterPro: IPR002965; P-rich_extensn.
 DR PRINTS: PRO1216; ADHESIVEI.
 DR PRINTS: PRO1217; PRICHEXTENS.
 KW Signal; Repeat; Hydroxylation.
 FT SIGNAL 1 20
 FT CHAIN 21 872 POTENTIAL.
 FT DOMAIN 21 41 ADHESIVE PLAQUE MATRIX PROTEIN.
 FT DOMAIN 124 872 NONREPEPTIVE LINKER.
 FT DOMAIN 184 192 TANDEN REPEATS OF Y-K-[PS]-K-[IP]-[ST]-Y-
 FT DOMAIN 213 221 P-[PSI]-[STI].
 FT DOMAIN 184 192 NONAPEPTIDE 1.
 FT DOMAIN 213 221 NONAPEPTIDE 2.
 SQ SEQUENCE 872 AA; 101677 MW; 98CC70D7C75FF3C4 CRC64;

DR EMBL: X63564; CAA5125.1; -
 DR EMBL: X74874; CAA52862.1; -
 DR EMBL: X74873; CAA52862.1; JOINED.
 DR EMBL: X74872; CAA52862.1; JOINED.
 DR EMBL: X74871; CAA52862.1; JOINED.
 DR EMBL: X74870; CAA52862.1; JOINED.
 DR PIR: S21054; S21054.
 DR MIM: 180660; -
 DR InterPro: IPR000684; RNA_polII_repeat.
 DR InterPro: IPR000722; RNA_pol_A.
 DR InterPro: IPR002879; RNA_pol_A2.
 DR Pfam: PF006623; RNA_pol_A; 1.
 DR Pfam: PF01854; RNA_pol_A2; 1.
 DR PROSITE: PS00115; RNA_pol_II_REPEAT; 43.
 DR Transferrase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
 KW DNA-binding; Nuclear protein; C2H2-type (POTENTIAL)
 FT ZN FING 71 87 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
 FT DOMAIN 1590 1958 W -> L (IN REF. 2).
 FT CONFLICT 1067 1067 D -> Y (IN REF. 2).
 FT CONFLICT 1449 1449
 SQ SEQUENCE 1970 AA; 217205 MW; 6876FC25692A657E CRC64;

Query Match 6.9%; Score 503.5; DB 1; Length 1970;
 Best Local Similarity 33.3%; Pred. No. 2,8e-13;
 Matches 161; Conservative 101; Mismatches 167; Indels 95; Gaps 47;

QY 307 PRTTPEKAPATTPKE-PRTTITKAPATTPKEPAATTTKSAPTTKE-----PAATTTKE 360
 DB 1507 PSMGSGSPMTPMNCATATAYGAMSPSVSGMTGCAAGSPASASGSPSPYSAMS 1566
 QY 361 PAATTPKEAPATTTKEPAATTTKSAPTTKEPAATTPKPAATTPKEP--APTTPKEPTP 418
 DB 1567 PTPGSPGSPGSPYIPSPGAMSPSPSPSPA-YEPRSPGVTGPQSPSPSPSPSP-SP 1624
 QY 419 TTPKEAPATTTKEPAATTTKSAPTTKEPAATTPKPAATTPKEPAATTTKEPSPPTTKE 478
 DB 1625 TTSPTSP-SPTSPNVSPTSPSY-SPTSPSY-SPTSPSY-----SPTSPSY 1671
 QY 479 PAATTTKSAPTTKEPAATTTKSAPTTKEPSPPTTKEPAATTTKEPAATTPKPAATTP 538
 DB 1672 -SPTSPSPSPSPSY-SPTSPSPSPSPSY-SPTSPSY-SPTSPSY-SPTSPSP-SP 1725
 QY 539 KEPAATTPKPAATTTKSAPTTKEPAATTPKPAATTPKPAATTPKPAATTPKPA 598
 DB 1726 SY-SPTSPSY-SPTSPSY-SPTSPNVT-PTSP-SYSPSPSP-SYSPSPNVT-T 1777
 QY 599 PTTPPELAPTTPEPTTTEEPATTPKPAATTPKPAATTPKPAATTPKPAATTPKPA 656
 DB 1778 PTSP-SPTSPSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSP 1828
 QY 657 TPKETAPTTTPKGAATTTKPAATTPK-KPAKELAPTTTKEPTSTSDKPAATTPKGA 715
 DB 1829 SP-SYSPSPSPSY-----PTSPSPSPSPSPSYPTTSPK--YSPTSPSPSPSPK-YS 1875
 QY 716 PTTPKKEAPATTPKPAATTTKGAATTTKPAATTPKPAATTPKPAATTPKPAATTPK 774
 DB 1876 PTSPPTV-SPTTPK-SPTSP--TYSPTSPVYTPPTSPK-----YSPSP--PTSPSPSP 1922
 QY 775 PAATTTKEAPATTPKKE--APTTPKPAATTPPTTSEVPTTKEPTTHKSPDE 832
 DB 1923 YSPSPSP-SYSPSPSPSPSYSPSPSY-SPTSP-----YSPSPSPSPSPSPSPSP 1965
 QY 833 STEPE 836
 DB 1966 SDEE 1969
 RESULT 14
 RPBL_MOUSE
 ID RPBL_MOUSE STANDARD; PRT; 1970 AA.
 AC P08775;
 DT 01-NOV-1988 (Rel. 09, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPBL).
 GN POLR2A OR RPO2-1 OR RPII215.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=67280135; PubMed=3038894;
 RA Aleatun J.M. Jr., Bartolomei M.S., West M.L., Cisek L.J., Corden J.L.,
 RT Cloning and sequence analysis of the mouse genomic locus encoding
 RT the largest subunit of RNA polymerase II.";
 RL J. Biol. Chem. 262:10695-10705(1987).
 RN [2]
 RP SEQUENCE OF 1567-1970 FROM N.A.
 RX MEDLINE=86068017; PubMed=2999785;
 RA Corden J.L., Cadena D.L., Aleatun J.M. Jr., Dahmus M.E.,
 RT "A unique structure at the carboxyl terminus of the largest subunit
 RT of eukaryotic RNA polymerase II.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:7934-7938(1985).
 RN [3]
 RP REVISIONS, AND PRESENCE OF AN ADDITIONAL EXON.
 RX MEDLINE=92178992; PubMed=1542581;
 RA Wintzerlth M., Acker J., Vicaire S., Vigneron M., Kedinger C.;
 RT "Complete sequence of the human RNA polymerase II largest subunit";
 RL Nucleic Acids Res. 20:910-910(1992).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
 CC RNA(N).
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
 CC -1- THE PHOSPHORYLATION ACTIVATES POL2.
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
 CC PRECURSOR, POLYMERASE II FOR THE MNNA PRECURSOR, AND POLYMERASE
 CC III FOR 5S AND TRNA GENES.
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: M12130; AAA40071.1; -
 DR EMBL: M14101; AAA40071.1; JOINED.
 DR PIR: A28490; A28490.
 DR MGD: MGI:98086; Rpo2-1.
 DR InterPro: IPR000684; RNA_polII_repeat.
 DR InterPro: IPR000722; RNA_pol_A.
 DR InterPro: IPR002879; RNA_pol_A2.
 DR Pfam: PF006623; RNA_pol_A; 1.
 DR Pfam: PF01854; RNA_pol_A2; 1.
 DR PROSITE: PS00115; RNA_pol_II_REPEAT; 42.
 KW Transferrase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
 KW DNA-binding; Nuclear protein; C2H2-type (POTENTIAL)
 FT ZN FING 71 87 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
 FT DOMAIN 1590 1958 P -> R (IN REF. 1 AND 2).
 FT CONFLICT 1498 1498 MISSING (IN REF. 1 AND 2).
 FT CONFLICT 1499 1536
 SQ SEQUENCE 1970 AA; 217175 MW; 7D76F38FD92A657E CRC64;

Query Match 6.8%; Score 498.5; DB 1; Length 1970;
 Best Local Similarity 33.1%; Pred. No. 4,3e-13;
 Matches 180; Conservative 101; Mismatches 168; Indels 95; Gaps 47;

```

QY 307 PPTPKKEPASTPKKE-PPTTIKSAPTTPKKEPAPTTTKSAPTPKKE-----PAPTTTKE 360
DB 1507 PSPGIGISPMATPNQCATPYGAMSSPVSGMTGGAAGSGSPASASGSPISYAWMS 1566
QY 361 PAPTTPKEPAPTTTKEPAPTTTTSAPTTTPKEPAPTTTPKRPAPTTTPKRP--APTTPKEPTP 418
DB 1567 PTPGSPGSPGSSPIYIPSPGAMSPSYSPSPSA-YEPRSPGQYRQSPSYSPSPSP--SP 1624
QY 419 TTPKEPAPTTKEPAPTTTPKEPAPPAKPPAPTTTPKEPAPTTTPKRPAPTTTKEPSPPTPK 478
DB 1625 TSPSY-SPTSPSYSPSPSY-SPTSPSY-SPTSPSY-SPTSPSY-----SPTSPSY 1671
QY 479 PAPTTTSAPTTPKEPAPTTTTSAPTTTPKEPSPPTTPKEPAPTTTPKRPAPTTTP 538
DB 1672 -SPTSPSYSPSPSY-SPTSPSYSPSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSP 1725
QY 539 KEPAPTTPKEPAPTTTPKRPAPTAPEPAPTTTPKEPAPTTTPKRLPTTPKEPAPTTPEKPA 598
DB 1726 SY-SPTSPSY-SPTSPSY-SPTSPNY-SPTSPNY-PTSP-SYSTSP-SYSTSPNY-T 1777
QY 599 PTPPEELAPTTPEEPTTPPEEAPTTTPKAAAPNTPEKAPTTTPKEPAPTTTPKRP--APT 656
DB 1778 PTSP-NKSPSPSY-SPTSPSY-SPTSP-SYSPSSPR---YTPQSP-YPSSSPSYSPS 1828
QY 657 TPKETAPTTTPKAPTTTPKKEPAPTTPK-KPAKELAPTTTPKEPSTSDKAPTTTPGTA 715
DB 1829 SP-SYSTSPSY-----PTSPSYSPSPSYTPTPASPK--YSPSPSPKSPSPK-Y 1875
QY 716 PTPKEPAPTTTPKEPAPTTTPKGAPTTPKKEPAPTTTPKRPAPKELAPTTTPKGP-TTSDK 774
DB 1876 PTSTPT-SPTTPK-SPTSP-SPTSP-SYSTSPSYTPTPSPK-----YSPSP--PTSPSPSPK 1922
QY 775 PAPTTPKEPAPTTTPKRP--APTTPKRPAPTTTPETTPPTSPVSTPTTPKEPTTHKSPDE 832
DB 1923 YSPSP--TSPSPSKGTYSPSPSY-SPTSP-----TSLTSPAI-----SPD 1965
QY 833 STPE 836
DB 1966 SDEE 1969

```

```

RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.
RX MEDLINE=87080760; PubMed=2878828;
RA Robinson P.A., Wion D., Anderson B.H.;
RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide
  (NF-H).";
RL FEBS Lett. 209:203-205(1986).
RN [4]
RX MEDLINE=89184647; PubMed=2928342;
RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D.,
  Smulowitz M., Carroll Z., Emanuel B.S., Breitner J., Rudin L.;
RT "Cloning of a cDNA encoding the rat high molecular weight
  neurofilament peptide (NF-H): developmental and tissue expression in
  the rat, and mapping of its human homologue to chromosomes 1 and
  22.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989)
CC CC
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
  AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NE-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
  SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -1- PHOSPHORYLATED ON A NUMBER OF REPEATS OF THE TRIPETIDE K-S-P. NFH IS
  THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
  INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
  OF AXONAL CALIBER.
CC -1- OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
  LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
  COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
  ONWARD AND IS LONGER DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation-
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M37227; AAA41693.1; ALT_FRAME.
DR EMBL: X13804; CAA32038.1; ALT_FRAME.
DR EMBL: M21964; AAA41695.1; -.
DR EMBL: J04517; AAA41692.1; -.
DR PIR: A30796; A30796.
DR PIR: A25649; A25649.
DR PIR: B25649; B25649.
DR PIR: S02003; S02003.
DR InterPro: IPR001664; IF.
DR Pfam: PF00038; filament; 1.
DR PROSITE: PS00226; IF; 1.
DR Intermediate filament; Coiled coil; Neurexin; Phosphorylation; Repeat.
KW NON_TER
FT DOMAIN 1 276 641
FT CONFLICT 164 164 51 X 3 AA TANDEM REPEATS OF K-S-P.
FT CONFLICT 185 185 L -> I (IN REF. 2).
FT CONFLICT 193 193 I -> S (IN REF. 2).
FT CONFLICT 199 199 L -> T (IN REF. 2).
FT CONFLICT 199 199 M -> T (IN REF. 2).
FT CONFLICT 346 346 K -> N (IN REF. 2).
FT CONFLICT 373 373 A -> V (IN REF. 1 AND 4).
FT CONFLICT 482 482 G -> E (IN REF. 2 AND 4).
FT CONFLICT 485 485 P -> S (IN REF. 2).
FT CONFLICT 570 571 RK -> KE (IN REF. 2 AND 4).
FT CONFLICT 591 591 P -> T (IN REF. 2 AND 4).
FT CONFLICT 727 727 A -> V (IN REF. 4).
FT CONFLICT 757 759 AAP -> GST (IN REF. 4).
FT CONFLICT 769 769 T -> L (IN REF. 2).
FT CONFLICT 775 775 R -> P (IN REF. 2 AND 4).
SQ SEQUENCE 831 AA; 89486 MW; 1B0973C3F13EF768 CRC64;

```

Query Match

6.8%; Score 498; DB 1; Length 831;

[illegible]

Search completed: April 26, 2002, 16:33:14
Job time: 624 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:26:07 ; Search time 61.21 Seconds
(without alignments)
500.359 Million cell updates/sec

Title: AA7
Perfect score: 7323
Sequence: 1 MANKTLPYLLLSVFEVIO.....ARAITRSGQTLSKVWVNCNP 1361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues
Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA: *
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PTCUS_COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	522	7.1	805	US-09-103-429A-4	Sequence 4, Appl
2	510.5	7.0	1837	US-08-928-361B-5	Sequence 5, Appl
3	506.5	6.9	744	5202236-25	Patent No. 5202236
4	498.5	6.8	786	US-09-103-429A-3	Sequence 3, Appl
5	496.5	6.8	1721	US-08-700-651-5	Sequence 5, Appl
6	496.5	6.8	1721	US-08-928-361B-6	Sequence 5, Appl
7	488.5	6.7	826	US-07-638-431-2	Sequence 2, Appl
8	488.5	6.7	826	PCT-US92-00018-2	Sequence 2, Appl
9	476.5	6.5	652	5202236-13	Patent No. 5202236
10	455.5	6.2	960	US-09-219-849-5	Sequence 5, Appl
11	428	5.8	1185	US-09-041-886-23	Sequence 23, Appl
12	424.5	5.8	1867	US-08-479-537A-5	Sequence 5, Appl
13	424.5	5.8	1867	US-09-083-116-5	Sequence 5, Appl
14	424.5	5.8	2035	US-08-479-537A-2	Sequence 2, Appl
15	424.5	5.8	2035	US-09-083-116-2	Sequence 2, Appl
16	420.5	5.7	907	US-08-783-774-2	Sequence 2, Appl
17	420.5	5.7	907	PCT-US95-04611A-19	Sequence 19, Appl
18	419.5	5.7	2476	US-08-276-967-2	Sequence 2, Appl
19	417	5.7	829	US-08-642-255-132	Sequence 132, App
20	417	5.7	829	US-08-397-633A-53	Sequence 53, Appl
21	417	5.7	837	US-08-175-155-68	Sequence 68, Appl
22	417	5.7	837	US-08-477-509B-103	Sequence 103, App
23	417	5.7	837	US-08-642-255-101	Sequence 101, App
24	417	5.7	837	US-08-707-237A-75	Sequence 75, Appl
25	417	5.7	837	US-08-482-085B-103	Sequence 103, App
26	417	5.7	897	US-08-397-633A-50	Sequence 50, Appl
27	404.5	5.5	408	US-07-609-716-65	Sequence 65, Appl

28	404.5	5.5	408	4	US-08-475-411A-65	Sequence 65, Appl
29	404.5	5.5	408	4	US-08-478-029A-65	Sequence 65, Appl
30	398.5	5.4	682	1	US-08-642-255-126	Sequence 126, App
31	398.5	5.4	682	1	US-08-397-633A-36	Sequence 36, Appl
32	385	5.3	1231	3	US-08-904-263A-4	Sequence 4, Appl
33	381.5	5.2	1537	1	US-08-325-267A-2	Sequence 2, Appl
34	381	5.2	1848	4	US-08-296-791-6	Sequence 6, Appl
35	381	5.2	1848	5	PCT-US95-10661A-6	Sequence 6, Appl
36	379.5	5.2	1187	1	US-08-320-559-28	Sequence 28, Appl
37	379.5	5.2	1187	3	US-08-545-860D-28	Sequence 28, Appl
38	379.5	5.2	1187	3	PCT-US94-04496-28	Sequence 28, Appl
39	379.5	5.2	1210	1	US-08-320-559-26	Sequence 26, Appl
40	379.5	5.2	1210	3	US-08-545-860D-26	Sequence 26, Appl
41	379.5	5.2	1210	5	PCT-US94-04496-26	Sequence 26, Appl
42	376	5.1	960	4	US-09-219-849-6	Sequence 6, Appl
43	371	5.1	761	2	US-08-707-237A-84	Sequence 84, Appl
44	371	5.1	762	1	US-08-642-255-114	Sequence 114, App
45	371	5.1	762	1	US-08-397-633A-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-09-103-429A-4
; Sequence 4, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang, Ping
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
; TITLE OF INVENTION: CDNA and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
; STREET: 118 No. 6187558th Tioga
; CITY: Ithaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,429A
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-2000
; TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 805 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; TISSUE TYPE: peritrophic membrane
; US-09-103-429A-4

Query Match 7.1%; Score 522; DB 4; Length 805;
Best Local Similarity 27.8%; Pred. No. 3e-25;

```
Matches 226; Conservative 33; Mismatches 211; Indels 342; Gaps 40;
QY 281 TSKYLAKPTPKAETTTGKPAITTPKEPTTPKPEASTTPKEPTPTTIKSAP-----TTP 335
Db 9 TALGLVAARPEVSDAEKNALHPHDPXP--PAQXXLLPXEYDCTKFYCEYGLKFIAP 66
QY 336 KEPAPTT-----TKSAPTTPKEPATTTTPKEPATTPKEPATTPKEPATTTK 383
Db 67 RDCAPGTEFEKSAQTCVHAALAGCTLCPGPAETT--QAPATTO--APTITQ--APTIT 119
QY 384 SAPTTTPKEPATTPKKEPATTPKEPATTPKEPATTPKEPATTPKEPATTPKEPAT 443
Db 120 QAPTTTQATTT-----QAPTTTQAPTTT-----QATTTQAPTTT----- 156
QY 444 PKPATTPTPKEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPAT 503
Db 157 ---QAPTITQ--APTITQ--APTITQ-----APTITQ--APTITQ--APTITQAA 198
QY 504 TTPKEPTTPKPEPATTP--KEPATTPPKKAPATTPKPEPATTPK----- 547
Db 199 TTPAATTPAAT--TPAATTPAATTPAATTPGVPAPTS---APVMPICELLNGCPADF 253
QY 548 -----PA-----BPAPTTK 555
Db 254 IHLPHDKYCNLFYQCSNGYTFEQRCEGLYFNYPVORCDSPANVECDGEISPAPVTE 313
QY 556 K-----PA----- 558
Db 314 GNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYQCVHGNLVERRCAGTHFSFEL 373
QY 559 -----PTAP--KEPAT-----TPKET 573
Db 374 QCDHIELVGCTLPGGESEEDVDDEDACTGWYCTEPIEWEPLNGCPADFSDHLLPHES 433
QY 574 -----APTTPKKLPPT-----PEKLAPTTPEKAP 599
Db 434 DCGYOLOCVHGQIARPCGNLHFSNATQSCSEPVTAGCVFECDSNQCTSTAAPTAAP 493
QY 600 TTPPELAPTPPEPTTPPEPATTPKAAAPNTPKPEPATTPKPEPATTPKPEPATTPK 659
Db 494 TAAPTAAPTAAPTAAPSTVVPPA--TPPATAAAPVPTTAIPT-----PAPTAAPTAAP 548
QY 660 ETAPTT---PKGTAPITLKEPATTPKKAPKELAPITTPKEPTSTSDKAPITPKGTAP 716
Db 549 PESPTTVVPTTAAPTA-----APT-----AVPEIPTVTSAPTAAPT--AAPTAAPTAAP 598
QY 717 TTPKEPATTPKPEPATTPKGTAPITLKEPATTPKKAPKELAPITTPKGTST--TSDKP 775
Db 599 TTAVPEIPTT-----VTSPTTAAPT--AAPAPT-----TVVPTTAAPTAA 641
QY 776 APTTPKETAPTTPKEPATTPKKEPATTPPTTPPTTSEVSTPTTTPKEPTTIHKSDE 835
Db 642 AP---NTTVTAPTAAPT--AAPAPNTTVVPTTAAPTAA-----PTVAH-----AP 685
QY 836 ELSAETPKALENSPKPGVPTTKTAATKPE 867
Db 686 NTTAAP-----VTTTSAPAIT--PE 703

RESULT 2
US-08-928-361B-5
; Sequence 5, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
```

```
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1837 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-928-361B-5
```

Query Match 7.0%; Score 510.5; DB 3; Length 1837;

Best Local Similarity 20.8%; Pred. No. 4.1e-24; Indels 559; Gaps 52;

Matches 314; Conservative 109; Mismatches 530;

```
QY 131 STTKRSPKPKPKKKT-----KKVIESEITEVKDNKKNRTKKPTPKPPVVDKAGS 181
Db 174 SITSGELKDPNKQATISGSRGCKGKQYSDSSTGRVDSITGLPTDPYPNCPNPVTGN 233
QY 182 GLDNGDFKVTTPDSTOHNKVVSTSPKITTAKP-----INPRSLPNSDTSKE--- 230
Db 234 LVSRSTGK--TIPNT---YAGVYRSNETKTPEPSANTNLLVDPKINAPCNSENSEPQV 288
QY 231 -----TSLTVNKETTVEKETTNTKQTSDDGKEKTTSAKETOSIEKTSADKLAPTSKVL 285
Db 289 IFDMGSKVPIPYTKCVGKHITTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 348
QY 286 AKPTPKAETTTGKPAITTPKPEPTTPKPEASTTPKEPTPTTIKSAPTTPKPEPAT 345
Db 349 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 408
QY 346 APTTPKEPATTTTPKEPATTPKPEPATTTTPKPEPATTTKSAAPTTPKPEPATTPK 405
Db 409 TTTTTTTTTTTTTTKKPTT-----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTKPTT- 460
QY 406 KEPATTPKPEPTTPKPEPATTPKPEPATTPKPEPATTPKPKKAPATTPKPEPAT 465
Db 461 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTATTTATTTKKPT--TTTTTTTTTTKKPTT 514
QY 466 TTKKPSPTT----- 475
Db 515 TTTATTTTTTSETSEVIKPDENWCLEKNGCEAKGATYGVIGDKGRIENGMAFTWIPND 574
QY 476 ----- 475
Db 575 DTHVRFKVKDVGNTISVRCGKAGKLEFPDRSLDFTIPPVAGHNSCSIIIVGVSOGGI 634
QY 476 -----PKEPAP-----TTTKSAPTPTTK 492
Db 635 HVSPYSGKDVSLISAPIQCELFENEYCDCTAKYGAIHSGYOTSADFVTTTTAKPTTTT 694
```


aa7.rai

Mon Apr 29 08:35:23 2002

```
QY 896 KETATTEKTESKITATTI 915
Db 720 YKAKPTYPSTYKAKPTYPST 739

RESULT 4
US-09-103-429A-3
; Sequence 3, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang, Ping
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
; TITLE OF INVENTION: CDNA and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
; STREET: 118 No. 6187558th Tloaga
; CITY: Ithaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103.429A
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-2000
; TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 786 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; TISSUE TYPE: peritrophic membrane
; US-09-103-429A-3

Query Match 6.8%; Score 498.5; DB 4; Length 786;
Best Local Similarity 29.1%; Pred. No. 8.8e-24;
Matches 200; Conservative 25; Mismatches 207; Indels 255; Gaps 32;

QY 288 PTAKAETTKGPAETTPKEPTTPKEPASTTPKEPTTIKSAPTTKEPAPTITKSAP 347
Db 93 PGPPAE-TTQAPAT-----QAPTTTQAPTTT-----QAPTTTQ--ATTITQAP 135
QY 348 TTPKEPAPTITKEPAPTITKEPAPTITKSAPTTPKEPAPTTPKAPPTTPKE 407
Db 136 TTTQ--APTTQ--ATT-----QAPTTTQ--APTTTQAPTTTQ--APTTQ-- 183
QY 408 PAPTTPKEPTTPKEPAPTITKEPAPTITKEPAPTAPKAPAPPTTPKEPAPT--PKEPA-- 464
Db 184 -APTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPA 241
QY 465 -----PTTTPKEPSPTPK-- 477
Db 242 ELLPNCGPADFIDHLLPHDKYCNLFYQCSNGVTFFQRCPEGLYFNPYVQRCDSANVEC 301

QY 478 ----EPAPTTT----- 484
Db 302 DGEISPAPPVTEGNEDEDIDGDLNDGCPANFEIDWLLPHGNRCDKYYQCVRHNLVERR 361
QY 485 -----KSAPTTTKEPAPTITTKSAPTTPKEPSPTTTPKE 517
Db 362 CGAGTHFSFELQOCDHIELVGLTLPGESEEDVDDEDA-CTGWYCTPEPIEWELPNGCP 420
QY 518 APTT-----PKEP-----APTTPKKAPTT----- 537
Db 421 ADFSIDHLLPHESDCGYLQCVHGQTIARPCFNLHFSPATQSCSPVTAGQVFECDSD 480
QY 538 ----PKEPAPTTPKEPAPTITTKKPAAPKAPAPPTTPKETAPTTPKKTTPPTPEKLA 594
Db 481 NQCTSTAAPTAAPTAAPTAAPTA--APSTVVPAPTPATAAPVPTT---AAPT-- 533
QY 595 EKPAPTTPEELAPTTPEEPTTPPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 652
Db 534 --PAPTAAPTAAPTAAPESTTVTVP--PTAAPTAAPTTAVPEIPTVT---SAPTAAPT 587
QY 653 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTTKEPTSTTSKPAPTTPK 712
Db 588 AAPTAAPTAAPTTAVPEIPTTVTSPTTAAPTAAPT--APNTT-----VTVP 632
QY 713 GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTTKEPTSTTS 772
Db 633 TAAPT--AAPANTTVTVPTTAAPTAAPTVAH-----APNTTAAPVTTTS 677
QY 773 DKAPTTPKETAPTTPKEPAPTTPKPK 799
Db 678 ---APATTPEDDDIDP--PLPNDP 699

RESULT 5
US-08-700-651-5
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700.651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
; US-08-700-651-5

Query Match 6.8%; Score 496.5; DB 3; Length 1721;
Best Local Similarity 22.1%; Pred. No. 2.9e-23;
Matches 295; Conservative 86; Mismatches 495; Indels 459; Gaps 43;

QY 110 PTPSPSSKAPPPSGASQTIKSTTKRSPKPPNKKTKKVKIESEITEVKDNKKNKTKKK 169
Db 82 PTPDPSNCPFNVTQ-NLVSRTSTGTIEN-----TYAGVYRSNETK--- 121
QY 170 TPKPPVVDAGSLONGDFKVTTPDTSTQHNKSVTSKITTAKPINRPSLPNDSKTSK 229
Db 122 TTEPSANTYAGVYRSN-ETKTTPEPSANT---NFLVDPKI-----NAPCSENSE 167
QY 230 E-----TSLTVNKEVTVEKETTITNTKQSTDGKEKTTSAKETQSIKTS 274
```



```
QY 1079 ---MDYLPRVNOGLIINP-----MLSDETNICMGK-----1106
Db 1207 PVDVTGLPYDPDSGEIIDPATKLPICGSVAGDEILVELNITDVTGLPIDLETGLPR 1266
QY 1107 -PVDGLTTLRNGTLV 1120
Db 1267 DVSGLPQLPNGTLV 1281

RESULT 7
US-07-638-431-2
; Sequence 2, Application US/07638431
; Patent No. 5198535
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khumsmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; TITLE OF INVENTION: immunogen and gene
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMRDC Building 1 T-12 National Naval
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/638,431
; FILING DATE: 19910110
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avrom D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-638-431-2

Query Match 6.7%; Score 488.5; DB 1; Length 826;
Best Local Similarity 26.4%; Pred. No. 3.9e-23;
Matches 153; Conservative 54; Mismatches 223; Indels 149; Gaps 34;

QY 324 TPTTIKSAPTTPKEPA-PTTIKSAPTTPKEPA-PTTIKEP-----APTTPKEP-APTTPKE 376
Db 273 TPCKVRCQCPQIPVPIPNKIPKPSNPEEPVNDPNDPNDPNDPNDPNDPNDPNDPNDPNN 332
QY 377 PAPTTTKSAPTTPKEP-APTTPKKP-APTTPKEP-APTTPKEP-APTTPKEP-APTTPKEP 432
Db 333 PNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNHPK 392
QY 433 PTTTPKEPAPTAPK--KPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 490
Db 393 RNPKRPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNK 443
QY 491 TKEPAPTTPTKSAPTTPKEPSPPTTTTKEPAPTTPKEPA-PTTPKKP-APTTPKEPAPTTPKE 548
Db 444 PKNPNP----NEPSNPNKPNPN-----EPLNPNPNEPSNPNPNPNPNPNPNPNPNPNPN 492

275 AKDLAPTSKVLAKPTPKAETTKGPALETPKPEPTPTTPKEPASTTPKPEPTPTTIKSAPT 334
209 -----PTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 249
335 PKEPAPTTTKSAPTTPKEPAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAP 394
250 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPT 309
395 TTPKKPAPTTPKEPAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKE 454
310 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 355
455 PAPTTTPKEPAPTTPKESPTTPKPEAPTTPKSAPTTPKPEAPTTPKSAPTTPKE----- 508
356 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTSETSVIK 415
509 -----PSPIT----- 513
416 PDEWCWLEKNGCEAKGATYGVIGKDGRIENGMAFTMIPNDHVRFRKVKVDVGNWIS 475
514 -----TKEP-----APTTPKEPAPTTPKPP 533
476 VRCRGAGKLEFPORSLSDFTPPVAGHNSCSIIYVSGDGKIHVSPYSGKDSVLS----- 531
534 APTTPKE-----PAPTTPKEPAPTTPKPAKAPTAPKAPKAPTTPKAPTAP 575
532 APIQPSSELFNEVYCDTAKYCAIHSGYOTSADFVTITTAKTPTTT--TGAPGQPTTTTT 589
576 TTPKKLTPTTPPKLAPTTPKEPAPTPELAPTTPPEAPTTPPEAPTTPPEAPTTPKAAAPTTPK 635
590 GSPSKPTTTTAKATTT-----TTLNPIITTTTQKPTTT-----TT---KVPKPP 634
636 EPAPTTPKEPAPTTPKEPAPTTPKAPTTPKGTAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 691
635 IATTTTLKPIVTTTAKATTTTIVPTT-----TTTKRDEMTTTTTTLPLPDIGDIEIT 689
692 PTTTKEPTS-----TTSKPAPTTPKG--TAPTTPKEPAPTTPKPEAPTTP 735
690 PPIEKMLDKYRMIDYNSGILLSDNDEIPGSOAGQIADTNSLFPVQTHKSGLPLDP 749
736 KGTAPTTPKE-----PAPTTPKPAKELAPTTPKGTSTTSKPAAPTTPKAPTTPKAPTTPKEPA 792
750 MVGLPDPKSGNLVHPYTNQMSGLSVSLAAKNLTVDTDTYG--LPIDTLTGYPPLDPV 807
793 PTPPKKPAPTTPETPPPTTSEVSTP-----TTKEPTTIHKS----- 829
808 SLIPFN--PETGELFDPISDEIMNGTIAGIVSGISASESLLSQKSAALIDPATNVMVGEFG 865
830 -----PDEST-----PELSAETPKALENSPKPEGVPTTKT 860
866 GLINPATGVMI PGLFPGSEQTSPLEIDGGIIPPEVAANADKFKLSIP--PSVP----- 919
861 PAATKPEMTTAKDKTTER-----DLRTTETTTAAPKMTKETATTE 903
920 --ESIP-----KQKIDSISELMYDIESGRIGQVSKRPIGSIAGOLNPNKPTQTD 972
904 KTTESKITATT--TQVSTTQDTPPEKITTLKTLTLLAPKVTITTKTITTEIMNKP--EE 960
973 SVTGKPIDPTTGLFPNPTGHLNPNNTMDSFAGAYKAVSNGIKTDNVGLPVGEI 1032
961 TAKPKD-----RAINSKATTPKPKOKPAKPKPTSTKKPKTMP----- 998
1033 TGLPKDPGSDIPENSTTGLVLDSTGKPINNSTAGIVSGKPGLPEDENGSLFDPSTNL 1092
999 -----RVKPKPTTPPKMTS--TWPE-----LNPT 1022
1093 PIDGNOLVNPENTNSTVSGSTGTTKPGIIPVNGGVVPEDEAKDQADKDGGLIVPPT 1152
1023 SRIEAMLOTTTRENTPNSKLVENPKSEADAGAEGETPHMLLRHVFVETPD---- 1078
1153 NSINKDPVNTQYSNTTGN-----INP--ETGKVIPGSLPGSLNPSFNTPOQTDEITGK 1206
```

QY 549 PATTITTKKAPAPKEP-APTTPKETAPTTPKKLTPTTPEKLAPTTPKEPAPTTPPEELAP 607
 Db 493 PS-----NPNEPSNPNEPSNPNE-----PSNPK-----PSNPE--P 523
 QY 608 TPEEPTPTTPEAPATTAKAAPNTPKEP-APTTKKEPAPTTPKEPAPTTPKETAPTTP 666
 Db 524 SNPNE--PLNPNP-----SNPNEPSNPNEPSNPEE--PSNPK-----PSNPE----- 564
 QY 667 KGTAPTTLKAPATTTPKKAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTPKEPAPTTP 726
 Db 565 -----PSNPEPNPEE--PSNPK-----SNPEEPINPEELNPKEPSNPNEESN 606
 QY 727 PKEPAPTTPKGTAPTTLKAPATTTPKKAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTP 786
 Db 607 KPEPT-----NPESNPKEP-----INPEDNENPLIODEPIERNDSNVIP 649
 QY 787 TPKEP-----APTTPKKAPATTTPPTTSEVSTPTTKEPTTIHKSDESTPELSAE 840
 Db 650 LPIIPKGNNTPSNLPENPSSEVEYPRNDNGENSNNTMKSNNI---PNEPIPSPGDN 706
 QY 841 PTPKALENSPK-----EPGVPTTK 859
 Db 707 PYKGHEERIPKPHRSNDYVYDNNVNNKKNKDEPEIPNNE 745

RESULT 8

PCT-US92-00018-2
 ; Sequence 2, Application PC/TUS9200018
 ; GENERAL INFORMATION:
 ; APPLICANT: Hoffman, Stephen L.
 ; APPLICANT: Charoenvit, Yupin
 ; APPLICANT: Hedstrom, Richard
 ; APPLICANT: Khushmith, Srisin
 ; APPLICANT: Rogers IV, William O.
 ; TITLE OF INVENTION: Protective malaria sporozoite surface protein
 ; TITLE OF INVENTION: Immunogen and gene encoding
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: A. David Spevack
 ; STREET: NWRDC Building 1 T-12 National Naval
 ; STREET: Medical Center
 ; CITY: Bethesda
 ; STATE: MD
 ; COUNTRY: USA
 ; ZIP: 20814-5044
 ; COMPUTER READABLE FORM: disk
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.24
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US92/00018
 ; FILING DATE: 19920103
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Spevack, Avram D.
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 295-6759
 ; TELEFAX: (301) 295-4033
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 826 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US92-00018-2

Query Match 6.7%; Score 488.5; DB 5; Length 826;
 Best Local Similarity 26.4%; Pred. No. 3.9e-23;
 Matches 153; Conservative 54; Mismatches 223; Indels 149; Gaps 34;
 QY 324 TPTTIKSAPTTPKEPA-PTTTKSAPTTPKEPA-PTTTKEP-----APTTPKEP-APTTIKE 376

Db 273 TPKVRCQPIPIPPVIPNKKIPEKPSNPEEPVNDNDNPNPNPNPNPNPNPNPNPNPNPN 332
 QY 377 PAPTITKSAPTTPKEP-APTTPKKP-APTTPKEP-APTTPKEP-APTTPKEP-APTTPKEP 432
 Db 333 PNNPN 392
 QY 433 PTTTKEPAPTAPK--KPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 490
 Db 393 RRPKRNRNPKPKNKNPNKNPNKNPNKNPNKNPNKNPNKNPNKNPNKNPNKNPNKNPN 443
 QY 491 TKEPAPTITKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 548
 Db 444 PKNPNP-----NEPSNPKNPNP-----EPLNPNPNPNPNPNPNPNPNPNPNPN 492
 QY 549 PAPTITTKKAPAPKEP-APTTPKETAPTTPKKLTPTTPEKLAPTTPKEPAPTTPPEELAP 607
 Db 493 PS-----NPNEPSNPNEPSNPEE--PSNPK-----PSNPE--P 523
 QY 608 TPEEPTPTTPEAPATTTPKKAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTPKEPAPT 666
 Db 524 SNPNE--PLNPNP-----SNPNEPSNPNEPSNPEE--PSNPK-----PSNPE----- 564
 QY 667 KGTAPTTLKAPATTTPKKAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTPKEPAPT 726
 Db 565 -----PSNPEPNPEE--PSNPK-----SNPEEPINPEELNPKEPSNPNEESN 606
 QY 727 PKEPAPTTPKGTAPTTLKAPATTTPKKAPKELAPTTTKEPTSTSDKAPATTTPKGTAPT 786
 Db 607 KPEPT-----NPESNPKEP-----INPEDNENPLIODEPIERNDSNVIP 649
 QY 787 TPKEP-----APTTPKKAPATTTPPTTSEVSTPTTKEPTTIHKSDESTPELSAE 840
 Db 650 LPIIPKGNNTPSNLPENPSSEVEYPRNDNGENSNNTMKSNNI---PNEPIPSPGDN 706
 QY 841 PTPKALENSPK-----EPGVPTTK 859
 Db 707 PYKGHEERIPKPHRSNDYVYDNNVNNKKNKDEPEIPNNE 745

RESULT 9
 5202236-13
 ; Patent No. 5202236
 ; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
 ; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILPULA, DAVID
 ; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
 ; PROTEIN
 ; NUMBER OF SEQUENCES: 39
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/528,762
 ; FILING DATE: 25-MAY-1990
 ; APPLICATION NUMBER: 82,456
 ; FILING DATE: 07-AUG-1987
 ; APPLICATION NUMBER: 933,945
 ; FILING DATE: 24-NOV-1986
 ; APPLICATION NUMBER: 650,128
 ; FILING DATE: 13-SEP-1984
 ; SEQ ID NO:13
 ; LENGTH: 652
 5202236-13

Query Match 6.5%; Score 476.5; DB 6; Length 652;
 Best Local Similarity 29.7%; Pred. No. 1.7e-22;
 Matches 221; Conservative 71; Mismatches 284; Indels 169; Gaps 45;
 QY 207 PKITTAKEINPRPSLPNSDTSKETSITVNKETTVEKTTTNNKQSTOCKEKTSAKE 266
 Db 11 PKMTYPTTKPKPSYPP-----TYKSKPTY-----KPKIT----- 40
 QY 267 TQSEKTSAKDLATPSKVLAKPTPKAETTTKGPALTTPKEPTTPPKE-----PASTTPKE 322
 Db 41 -----YPPTYK--AKPS-----YPPTYKPKKTYPTTKPKLTYPPTYKPKP 79

GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
TITLE OF INVENTION: Proapoptotic Peptides, Dependence
POLYPEPTIDES AND METHODS OF USE
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041.886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1185 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-041-886-23

Query Match 5.8%; Score 428; DB 4; Length 1185;
Best Local Similarity 22.1%; Pred. No. 3.7e-19;
Matches 237; Conservative 114; Mismatches 429; Indels 292; Gaps 50;

Qy 117 KKAPPP-----SGASQTKSTTKRSPKPPNKKTKKVVESIEITVEVKNNKRNKKPK 169
Db 18 KEAPGPREELSRASPGVSSSSDCKAEKSKQAKARVEAEASTPKYKQGRSEE-- 75
Qy 170 TPKEPVVDEAGSLNDGDFKVTTPDTSTQHNKYSTSPKITTAKPINRPSLPPNSDTSK 229
Db 76 -----ISESES-----BETNAPKTKTEQELPRQSPSDLDLSD 109
Qy 230 ETSLVNKEITVETKETTNNKOTSDGKEKTSKAKETOSIEKTSKADLAPTSTKSKAP 289
Db 110 GRSL--NDGSSDRIDQDNRS-----PSIYSPGSVENDS----- 145
Qy 290 PKAETTKGPAITPKPTTP--KEPASTTKEPTPTTIKSAPTTPKEPATTTKSAP 347
Db 146 DSSSGLSQGA--RPVHPPLFPSPQPDSTPRQEAFF-----EPHPSVPTGY 194
Qy 348 TTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTP 390
Db 195 HAPWER-PTSRMFPAP--PGAPPPHQLYPGGTGVLSPGMPKGGGAASSVGGNGCK 251
Qy 391 -EPATTTPKPPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATT 443
Db 252 QHPPTTPISSVSSGASGAPTKP-----PTTPVGGNLPSPAPPPANFHVTPNLPPPPAL 307
Qy 444 -PKPAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATT 494
Db 308 RPLNNASAP----PGLGAQPLPGHL--PSYAMGGMGGLPGP--EKGTLPSPHS 358
Qy 495 APPTTKSAPTTP-KEPSPTTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKE 552
Db 359 LPASSAPAPMWRPYSS 416

Qy 553 TTKKPAPTAPKEPATTTPKETAPTTPKKLT-----PTTPE--KLAPTTPEKAPTTPPEEL 605
Db 417 SL-----SVSNQP-----PKYTQPSLQAVWSQGGPPPPYGRLLANSNAHPGPPFSTG 467
Qy 606 APTTPEEPTPT-----TPEEPAP-----TTPKAA 629
Db 468 AOSTAHPVSTHHHHHQQOQQOQQOQQOQQOHHGNSGPPPCAFPHPLEGGSSHHAHYAM 527
Qy 630 APN-----TPKEPAPTTPKEP-----APTTPKEPAPTTPKETAPTTPKGPAPTTLKEPAP 679
Db 528 SPGLSLRPPPGPAHLPPPHSHQSVSYQAQGNPGPPVSSSSSSSSSSSSSSSSSSSSSSSSSS 585
Qy 680 TTPKKPAPKELAP---TTTKEPTSTTSDDKPAPTTPKCTAPTTPKEPAPTTPKEPAPTTPK 736
Db 586 SQGQGAPYPPFPVPTVTSSATLSTVIATVASSPAGYKTAASPPGPPYKGRAPSPGAYK 645
Qy 737 GTAPTTLKEPAPTTPKKPAPKELAPTTPKPTSTTSOKPAPTTPKETAPTTPKEPAPTTP 796
Db 646 TATPPGYKPGSP-----PSFTGTGPGYRG-----TSPAGPGTEKPGSPV--GPGPLPP 694
Qy 797 KKPAPTTPETPPPTTSEVSTPTTKEPTTIHKSPDE--STPELSAETPKALENS--PKE 852
Db 695 AGPS-GLPSLPPPPAPASGPPLS--ATQIKQEPAEYEYETPE---SPVPPARSPSPPKV 748
Qy 853 PGVPTTKTAAKPEMTTAKDKTTERDLRTTPTTAAKPKMTKETATTEKT-TESKIT 911
Db 749 VDVPSHASQAREFNKHLDRGNFSCARSDLYFVP---LEGSKLAKRADLVEKVRRAEOR 805
Qy 912 ATTTQVSTTTQDTPFKITTLKTTTLAPKVTITTKTITTTIMNKPDEETAKPKDRATNS 971
Db 806 AR-----EKEKEREREREK 820
Qy 972 KATTPKQKPTKAPKPTSTTKKPTMPVRKPKTTTPRKMSTMPELNPTSRIAEAMLO 1031
Db 821 EREKERELERSVKLAQEGRAP-----VECPSLGVPVPHR-----PPFPGSAVA----- 865
Qy 1032 TTPRNPQNTNSKLVEVNPKSEDAGGAEGTPHMLRPHVFMPEVTPDMDYLP 1083
Db 866 -TPPYLPGDTPALRT--LSEYA-----RPHVMSPGNRNHPFYP 902

RESULT 12
US-08-479-537A-5
Sequence 5, Application US/08479537A
Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMBERN, Pierre
APPLICANT: KIEN, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUVENI, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990

aa7.rai

Mon Apr 29 08:35:23 2002

PRIOR APPLICATION DATA: WO PCT/FR91/00835
APPLICATION NUMBER: 23-OCT-1991
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA: US 08/039,320
APPLICATION NUMBER: 04-APR-1993
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA: US 08/403,576
APPLICATION NUMBER: 14-MAR-1995
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note= "The amino acids spanning 128 to 1727 constitute a repeated region wherein the repeat is 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a 21 amino acid precursor sequence."
OTHER INFORMATION: 21 amino acid precursor sequence."

US-08-479-537A-5

Query Match 5.8%; Score 424.5; DB 2; Length 1867;
Best Local Similarity 23.7%; Pred. No. 1e-18;
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 231 TSLTVNKTETVETKTTTNTKSTGKETSATQSIKTSKAD-LAPTSKVLAKPT 289
DB 16 TVLTV-----VTGSHASSTPGGKETSATORSSVPSSSTKNAVMTSSVLSSH 65
QY 290 P-KAETTTKGP-ALITPKP-----tpttpkpeasttpkpepttttksap 332
DB 66 PGSSSTTQGDVTLAPATPAGSGAATWQDVTSVPVTPALGSTTPPAHDT----SAP 122
QY 333 TTPKEAPTPT-----KSAPITPKP-----APTTTKEAPTTPKEAPTPT 373
DB 123 --DNKPAGSTAPXAHGVTAPDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVT 180
QY 374 TKE--PAPTT-----KSAPITPKPAPTTPK-----KPAPTTPKEAPTTPKEPTPT 420

Db 181 APDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVT 240
QY 421 PKE--PAPTTKEP-----APTTTKEPAPTAPK-----KPAPTTPKEAPTTPKEAPTPT 467
Db 241 APDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVT 300
QY 468 TKEPSPTTPKEAPTPT--TKSAPITTTKEAPTPT-----TKSAPITTPKEPSPTTTK----E 516
Db 301 APDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVT 359
QY 517 PAPTTTPKEAPTTPK-----KPAPTTPKEAPTTPK-----EPAPTTPKAPAPTAPKEAPT 568
Db 360 SAPDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVT 419
QY 569 TPKETAP-----TTPKKLTPTTPEKLAPTTPEKAPTPEELAPTTPEEPTPTTPEAPT 624
Db 420 SAPDXRPPGSTAPXAHGVT-----APDXRPPGSTAPXAHGVTAPDXRPP-----PGST 471
QY 625 TPKA-----AAPNTPEAPTTPKEAPTTPKEAPTTPKAPTTPKGT--APTTLKEPAP 679
Db 472 APXAHGVTAPDXRPPGSTAPXAHGVTAPDXRPP--XPGSTAPXAHGVTAPDXRPPXGS 530
QY 680 TTP-----KKPAPKELAPTTTKEPTSTTSKAP--TTPKGTAPTTPKEAPTTP 727
Db 531 TAPXAHGVTAPDXRPPGSTAP--XAHGVTAPDXRPPGSTAPXAHGVT-----APDXR 585
QY 728 KEAPTTPKG---TAPTTLKEAPTTP-----KKPAPKELAPTTTKEPTSTTS 773
Db 586 PXPGSTAPXAHGVTAPDXRPPGSTAPXAHGVTAPDXRPPGSTAP--XAHGVTAPDX 644
QY 774 KPAP--TTPK-----ETAPTTPKEAPTTP-----KKPAPTTPETPTTSEVSTP 817
Db 645 RXPXGSTAPXAHGVTAPDXRPPGSTAPXAHGVTAPDXRPPXGS--TAPXAHGVTAP 702
QY 818 TTTKEPT-----THKSPDESTPELSAPTPKALENSEKPEGVPTTKTPAA----- 863
Db 703 DXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVTAPDXRPPGSTAPXAHGVTAP 762
QY 864 -TKPEMTTTAKDK---TTERDLRTPTTTPAAMKMKETATTTTEKTESKITATTQVTS 919
Db 763 DXRPPGSTAPXAHGVTAPDXRPPXPGST--APXAHGVTAPDXRPPGSTAPXAHGVT 820
QY 920 TTTQDTPPFKITTLTKTTLAPKVTITTKTITTTIMMKPEETAKPKDRATNSKATTPKQ 979
Db 821 APDXRPPXG-----STAPXAHGVT-----SAPDXRPPGSTAPXAHGVTAPDXRPP- 868
QY 980 KPTKAPKAPTSTKKPKTTPRVKPKTTTPPKMTSTMPKMTSTMPKMTSTMPKMTSTMPKMT 1039
Db 869 -GSTAPXAHGVTAPDXRPP--XPGSTAPXAHGVTSA-PDXRPP-----PGST 911
QY 1040 -PNSKLVENPKSEDAAGAGETPHMLLRPHVFMPEVTPDMXYLPVPNOGIIINPLSD 1098
Db 912 APXAHGVTAPDXRPPXPGSTAPXAH-----GVTAPDXRPPXPGSTA----PAXHG 957
QY 1099 ETNICNGKPPVDGLTT-LRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIPSPID 1153
Db 958 VTSAPDXRPPXPGSTAPXAHGVTSA-----PDXRPPXPGSTAPXAHGVTAPD 1003

RESULT 13
US-09-083-116-5
; Sequence 5, Application US/09083116
; Patent No. 6203795
; GENERAL INFORMATION:
; APPLICANT: CHAMON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L. L. P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/0083.116

OTHER INFORMATION: /note= "The amino acids spanning
 128 to 1727 constitute a repeated region wherein the repeat
 unit is 20 amino acids, 17 of which are fixed. The number of such
 repeats varies from 1 to 40."
 OTHER INFORMATION:

```

FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:

```

```

NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION:
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:

```

NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is x2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:

NAME/KEY: Peptide
LOCATION: 1, 21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
083-116-5

US-09-083-116-5

Db 703 DXRXPXGSTAPXAHGVTSPDXRXPXGSTAPXAHGVTSPDXRXPXGSTAPXAHGVTSP 762
Qy 864 -TKPEMTTAKDK---TTERDLRTTPTTAAAPKMTKETATTEKTESKITATTTQVTS 919
Db 763 DXRXPXGSTAPXAHGVTSPDXRXPXGST--APXAHGVTSPDXRXPXGSTAPXAHGVT 820
Qy 920 TTTQDTPPKITLTKTTTLLAPKVTTKKTTTITTEIMNKPEETAKPKDRATNSKATTPK 979
Db 821 APDXRXPXG-----STAPXAHGVT-----SAPDXRXPXGSTAPXAHGVTSPDXRXP 868
Qy 980 KPTKAPKPTSTKPKTMTBRVRPKTTPTRKWTSTMPLENTSRIAEAMLQTTTTRPQ 1039
Db 869 -GSTAPXAHGVTSPDXRXP--XPGSTAPXAHGVTSA-PDXRXP-----PGST 911
Qy 1040 -PNSKLVFNPKSDEAGGAEGETPHMLLRPHVFMPEVTPDMYLPVRPNQGIINPMLSD 1098
Db 912 APXAHGVTSPDXRXPXGSTAPXAH-----GVTSAPDXRXPXGSTA-----PXAHG 957
Qy 1099 ETNICKGKVDGLTT-LRNGTLVAFRGHYFWMLSPFSPARRITEVWGIPIPSID 1153
Db 958 VTSAPDXRXPXGSTAPXAHGVTSA-----PDXRXPXGSTAPXAHGVTSPDX 1003

RESULT 15
US-09-083-116-2
; Sequence 2, Application US/09083116
; Patent No. 6203795
; GENERAL INFORMATION:
; APPLICANT: CHAMBERON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,116
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/479,537
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
128 to 1899 constitute a repeated region wherein the repea
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 - Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro - CC
OTHER INFORMATION: CCC, CCA, or CCG; and Ala - GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y - Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr - ACT, ACC,
OTHER INFORMATION: or ACG; and Asn - AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 - Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro - CCT, CCC,
OTHER INFORMATION: or CCG; and Ala - GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-2

Query Match 5.8%; Score 424.5; DB 4; Length 2035;
Best Local Similarity 23.7%; Pred. No. 1.le-18;
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;
Qy 231 TSLTVNKTETVETKETTNTKQSTDKGKTSKAKTOSIEKTSKAD-LAPTSKVLAKPT 289
Db 16 TVLTV-----VTSGGHASSTPGGKEKTSATQSVSSTKNAVSTSSVLSHS 65
Qy 290 P-KAETTTKGP--ALTTPKEP-----TPTTPKEPASTTTPKEPTTIKSAP 332
Db 66 PGSGSTTGGQDVTLAPATEPASGAATWGQDVTSVPVTRPALGSTTTPPAHDVT---SAP 122
Qy 333 TTPKEPAPTTT-----KSAPTTTPKEP-----APTTPKEPAPTTTPKEPAPT 373
Db 123 --DNKPAPGSTAPXAHGVTSPDXRXPXGSTAPXAHGVTSPDXRXPXGSTAPXAHGVT 180
Qy 374 TKE--PAPTTT-----KSAPTTTPKEPAPTTPK-----KPAPTTTPKEPAPTTTP 420
Db 181 APDXRXPXGSTAPXAHGVTSPDXRXPXGSTAPXAHGVTSPDXRXPXGSTAPXAHGVT 240
Qy 421 PKE--PAPTTKEP-----APTTPKEPAPTPAK-----KPAPTTTPKEPAPTTTPKEPAPT 467
Db 241 APDXRXPXGSTAPXAHGVTSPDXRXPXGSTAPXAHGVTSPDXRXPXGSTAPXAHGVT 300
Qy 468 TKEPSPTTPKEPAPTT--TKSAPTTTTPKEPAPTT-----TKSAPTTTPKEPSPTTK----E 516
Db 301 APDXRXP--XPGSTAPXAHGVTSPDXRXPXGSTAPXAHGVTSPDXRXPXGSTAPXAHGVT 359
Qy 517 PAPTTTPKEPAPTTTPK-----KPAPTTTPKEPAPTTTPK-----EPAPTTTKKPA 568
Db 360 SAPDXRXPXGSTAPXAHGVTSPDXRXPXGSTAPXAHGVTSPDXRXPXGSTAPXAHGVT 419
Qy 569 TPKEPAP-----TTPKKLTPTTTPKEPAPTTTPKEPAPTTPELAPTTPEEPTTPPEPAPT 624
Db 420 SAPDXRXPXGSTAPXAHGVTSPDXRXPXGSTAPXAHGVTSPDXRXPXGSTAPXAHGVT 471

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:32:49 ; Search time 114.61 Seconds
(without alignments)
871.345 Million cell updates/sec

Title: AA8

Perfect score: 7064

Sequence: 1 MAWKTLPIYLLLSVEVIO.....ARAITRSGQTLSKVWVNCNP 1311

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

PIR_68:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1078.5	15.3	3020	2 A43932	mucin 2 precursor,
2	950	13.4	1664	2 T18262	S-layer protein -
3	855	12.1	1489	2 T31108	cyst germination s
4	802	11.4	1274	2 T16251	hypothetical prote
5	798.5	11.3	1367	1 S48478	glucan 1,4-alpha-g
6	784	11.1	1188	2 S49915	extensin-like prot
7	776.5	11.0	2187	2 T30826	nascent polypeptid
8	672	9.5	1344	1 A35175	mucin 1 precursor,
9	671	9.5	1151	2 T18535	high molecular mas
10	662	9.4	1229	2 T25697	hypothetical prote
11	661	9.4	3507	2 T34513	hypothetical prote
12	646.5	9.2	3570	2 T45025	hypothetical prote
13	635.5	9.0	7962	2 T38346	mucin MUC5B, trach
14	633	9.0	489	2 T11622	elastic titin - hu
15	632	8.9	761	2 C84672	extensin class 1 p
16	630.5	8.9	990	2 T1618	hypothetical prote
17	626.5	8.9	971	2 T19431	nucleolar phosphop
18	625	8.8	6642	2 T29757	hypothetical prote
19	622.5	8.8	839	2 F75518	protein UNC-89 - C
20	608	8.6	801	2 T29018	hypothetical prote
21	607.5	8.6	924	2 S27923	hypothetical prote
22	605	8.6	379	2 S50125	gene LF3 protein -
23	559.5	7.9	350	2 S22456	larval glue protei
24	556.5	7.9	856	2 T16543	hydroxyproline-ric
25	556.5	7.9	875	2 S23760	hypothetical prote
26	555.5	7.9	1630	2 A53577	polyphenolic adhes
27	554	7.8	2232	2 T34434	ascites sialoglyco
28	551	7.8	620	2 S06733	hypothetical prote
29	545	7.7	873	2 A47283	hydroxyproline-ric
					calphotin - fruit

30	542	7.7	369	2 S20500	hydroxyproline-ric
31	538.5	7.6	416	2 J00465	extensin precursor
32	533	7.5	1087	1 QFMSH	neurofilament trip
33	532	7.5	756	2 T27642	hypothetical prote
34	530.5	7.5	1162	2 JH0557	exo-alpha-sialidas
35	530	7.5	865	2 A47282	calcium-binding pr
36	522	7.4	328	2 J00985	hydroxyproline-ric
37	522	7.4	1459	2 T32271	hypothetical prote
38	518	7.3	813	2 S70795	vsA protein precu
39	518	7.3	866	2 T45462	membrane glycoprot
40	518	7.3	1072	1 A37221	neurofilament trip
41	512.5	7.3	867	2 T45463	membrane glycoprot
42	509.5	7.2	662	2 A45155	mucin FIM-C.1 - Af
43	508	7.2	1832	2 T31113	mucin-like glycopr
44	505	7.1	700	2 A54641	interspersed repea
45	504.5	7.1	606	2 A43427	neurofilament trip

ALIGNMENTS

RESULT 1

A43932

mucin 2 precursor, intestinal - human (fragments)

N:Alternate names: mucin SMUC-41

C:Species: Homo sapiens (man)

C>Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999

C:Accession: A49963; A45106; B45106; A43932; B33532; A61257; P00328; P00329

R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.

J. Biol. Chem. 269, 2440-2446, 1994

A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t

A:Reference number: A49963; MUID:94132002

A:Accession: A49963.

A:Molecule type: mRNA

A:Residues: 1-639 <GU1>

A:Cross-references: GB:L21998

R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.

J. Biol. Chem. 267, 21375-21383, 1992

A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up

A:Reference number: A45106; MUID:93016075

A:Accession: A45106

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 626-1895 <GU2>

A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396

A:Note: sequence extracted from NCBI backbone (NCBIP:116706)

A:Accession: B45106

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 2037-3020 <GU3>

A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398

A:Experimental source: colon

A:Note: sequence extracted from NCBI backbone (NCBIP:116698)

R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,

J. Clin. Invest. 88, 1005-1013, 1991

A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym

A:Reference number: A43932; MUID:91358717

A:Accession: A43932

A:Molecule type: DNA

A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>

A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864

A:Note: sequence inconsistent with the nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)

R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.

J. Biol. Chem. 264, 6480-6487, 1989

A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evl

A:Reference number: A33532; MUID:89197956

A:Accession: B33532

A:Molecule type: mRNA

A:Residues: 1916-2193 <GU4>

A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874

A:Experimental source: intestine

R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

QY	112	KNRTKK-KP-----TP-----KPPVDEAGSLDRNGD-----FKVTTPTDTSTQHNKVVSTS	156
Db	274	KNPTKKWPPWEDETPEVEBKPPPEKKAPVLKKDPAPAAKADPSPSKAAPKKVPS	333
QY	157	PKTITAKPI-NRPSPLPNSDTSKETSILNVNETTVETKETTITTKNOTSTDKCKETTISAK	215
Db	334	SPVVPTTPVKNPKYKPPWEVDDEPAEEVKPSAPEKKTPLVKRKEPSPSTTSSDPS	393
QY	216	ETOSIEKTSKADJAPTSKVLAQTPPKAE-----TTTKGPA-----LTP	254
Db	394	PKKAAPAVKPRDSSPKKATPLQADPKAQOVPPTPVKNPKYKPPWEVDDEDPVEEVKQP	453
QY	255	KEPTPTTP-----KEPASTTPKEP-----TP-TTTSAPTTTPKEPAPTTTKSAPTTTPKEPAP	305
Db	454	EAPAKKTPVLKRKEPAKADTAKTSKTPETPEKDPVKPROSSPKVAAKPDQAQAPA-	512
QY	306	TTTKGPA-----PPTPKAPAPITTKKEPAPT-----TTKSAPTPP	339
Db	513	TPVKNPKYKWRPWEDEETPADVDVKPTDAKKTPSLAKDPAPAKESLPKADTRAPAKP	572
QY	340	KEP-----APTTKPKPAPTTPE-----PAPTPKEBETPTTPKEPAPTTTPKEPAPTTPK-BP	389
Db	573	RDPSPKKVAPTAPEKKTTPVLAKKEPAGPADSKTKEPKSKPRDSPKKVAAKVPKTEV	632

RESULT 5

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (*Saccharomyces cerevisiae*)
S48478
N;Alternate names: extracellular glucosylase, mucin-like protein MUC1, pro
N;Species: *Saccharomyces cerevisiae*
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C;Accession: S48478; A26877; B26877; S27281; JC6123
R;Rowley, K.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48478
A;Accession: S48478
A;Molecule type: DNA
A;Residues: 1-1367 <ROW>
A;Cross-references: GB:247047; EMBL:Z38061; NID:9603997; PID:g763364; GSPDB:
R;Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A;Title: Gene fusion is a possible mechanism underlying the evolution of ST
A;Reference number: A91831; MUID:87194500
A;Accession: A26877
A;Molecule type: DNA
A;Residues: 1-242 <YAM>
A;Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
A;Accession: B26877

Db 601 AP-PAHGVTSAPDTRPAGSTAPPAHGVTS-----APDTRPAGSTAPPAHGVTSAPDTRP 655
QY 695 EPAPTTTP-----KKPAPKELAPTTTKGPTSTSDKPAP--TPK-----ETAPTTTP 738
Db 656 APGSTAPPAHGVTSAPDTRPAGSTAP-PAHGVTSAPDTRPAGSTAPPAHGVTSAPDTR 714
QY 739 KEPAPTTTP-----KKPAPTTTPETPTTSEVSTPTTKEPTT-----IHSPD 781
Db 715 PAPGSTAPPAHGVTSAPDTRPAGS--TAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPD 772
QY 782 ESTPELSAETPKALENSPKPEGVPTTKTPAA-----TKPEMTTAKDK-----TTERD 830
Db 773 TRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPD 832
QY 831 LRTTPEPTT-----TAAPKMTKETATTTTEKTTESKITATTTQVTTSTTODTTP--FKIT 881
Db 833 TRPAGSTAPPAHGVTSAP-----DTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGV 888
QY 882 TLKTTTTLAPKVTTTKTITTT--EIMNKPETAKPKDRATNSKATTPKPKQTKAPKKPT 939
Db 889 SAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAP--GSTAPPAHG 946
QY 940 STKKPKTMRPRVKKTTPTRPKMTSTMPELNP-----TSRIAEAMLQ--TTRP--NOTPN 991
Db 947 VTSAPDTRP--AFGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPP 1003
QY 992 SKLVEVNPKSEDAGAEGETPHMLLRPHVFEVTPDMDYLPRVPMOGIILNPMLSDET 1051
Db 1004 AHGVTSAPDTRPAGSTAPPAH-----GVTSAPDNRPALGSTA-----PPVHNVS 1049
QY 1052 ICNGKPVGLTLRLNGTLVAFRGHYFWMLSPFSPS 1087
Db 1050 ASGSAGSASTLVHNGTSARATTTTPASKSTPFSIPS 1085

RESULT 9
T18535
high molecular mass nuclear antigen - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18535
R:Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick
A:Reference number: Z18955; MUID:9803440
A:Accession: T18535
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1151 <SHI>
A:Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BAA24137.1

Query Match 9.5%; Score 671; DB 2; Length 1151;
Best Local Similarity 25.9%; Pred. No. 8.6e-22;
Matches 284; Conservative 100; Mismatches 419; Indels 292; Gaps 55;

QY 120 TPKP-----PVVDAGSLGNDGFKVTTPTDSTTQHNKVS--PKITTAKPINRPS 170
Db 113 TTPPSQGPAGTPPSQAGAPKGDGTQPSGTSKADCKPAAQDVPRATTA-ATEARP- 170
QY 171 LPPNSDTSKETSILVNKETTVEKETTNN--QTSDGKETTSAKETQSTKESAKDL 228
Db 171 -----ASAASPTVPKATAEATVTAASQSAPKAAIDAAAVTAA---SQSAPKATV-EV 219
QY 229 APTSKVLAKPTKAEITTKGPAITTPKEPTP-TTPKEPASTTPKEP---TPTTKTSAPT 283
Db 220 KPAATAVAKAEKAVTAAAAAPKATAEAKPAPVTSPTIPICSSAEAKPLTAASFTASKA--T 277
QY 284 TPKPAPTT-----TKSAPTTTPKEPAPTTTK----- 309
Db 278 AEAKVPATASLMATKVTAEAKPAPSPVPKATTDTKAVTATAPKAGPDVVKPAVAVCABA 337
QY 310 EPAPTTTPKEPAPTTTPKEPAPTTTKSAPTT-----PKPA-----PTTPKKPAPTTPK 356

Db 338 KPAPPPPOQLPKAAAAAAPTGTCLKPATAPPHGSPRANSHVTVTVPNVAAAAATVP- 396
QY 357 EPAPTTTPKEPPTT-----PKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 412
Db 397 -TAGAVPKRAGTGTTPAAAAPQOPVP---KAAPVTPPSQOAVPRAATAAA-----APVTPQ 448
QY 413 P---APTTPKESPTTPKEPAPTTTKSAPTTTKPAP---TTTKSAPTTTPKEPSPTTTKE 466
Db 449 PVTKAATTTNATPPQPIPKAATTTTATVPVPOQIPKAGTDAAPPPAVKRAPSDGRAAT 508
QY 467 P-----APTTPKAPAPTTTPKAPAPTTTPKAPAP-----TTPKAPAP-TTKKAPAP 512
Db 509 PGVPAATDPOKPPPTPQSPVSAVTEPKPQRAAPPPSNEATPAVPSPNLKSPLPTIP 568
QY 513 K-----EPAPTTK---ETAPTTPKLTPTTPSKLAPTTTPKEKAPAPTTPELA----- 556
Db 569 KPVPMLALTPOQVTAQMTQLAATKPSPIVKASPK-ALMTPPPPPPGLPRALAAAKLLG 627
QY 557 -PTTP-----EETPTTTP---EEPAPTTTP-----KAAAPNTPK----- 586
Db 628 LPSSPVASAMHAKVTPRPLPASVPVMAASPASILGPDARVALATNAASPGAKPEAAGNG 687
QY 587 ----PAPTTTPKEPAPTTTPKEPAPTTTPKETAPT-----TPKGTAPT----- 622
Db 688 TLMAPMAAANTOMAPIGAAGAAQTAPMGAAHTHVSPMGAGGATQMSPTGAANTHMSPIGA 747
QY 623 -----TLKEPAPTTTPKAPKELAPTTTKP-----TSTTSDKPAPTTTP 661
Db 748 GGATQMSPMGAANTQSPMGATTTQSPMGAAATQSPMGAAATQVATATAGNTMQVSP 807
QY 662 KG--TAPTTTPKEPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPK 717
Db 808 MGAATPPQTPSVGAATTP-QPSPM-----GAATTTLMSPMGAATTPQ---EPMGAVTTQPP 859
QY 718 ----TSTTSDKP-APTTPKET---APTTPKET---APTTPKET---APTTPKET--- 758
Db 860 PMAATNTQPPMAASTPQSTPMGAATTTQSPMGATTTQSPMGATTTQSPMGASTPQAPTVAGSPT 919
QY 759 PTTSEVSTPTTTPKEPTTIHKSDESTPEL--SAEPTPKALENSPKPEGVPTTKTPAATKP 816
Db 920 PPPPIPSPTAOTSPQMSKSPDPKAPSAATQSPAHHVANASPGV-TAVSPA---P 975
QY 817 EMTTAKDKTTERDLRTTPTTTTAAKMT-KETATTTETKTESKITATTTQVTTSTTODT 875
Db 976 IGVTEASPSADGARLSGPTAATDGPKASPAATADVTEAATD--VTAATAVPA----EA 1029
QY 876 TPKITTLTKTTTLAPKVTTTTKTITTTTINMKPEETAKPKDRATNSKATTPKPKPTKAP 935
Db 1030 AP-----TKAKRSS 1077
QY 936 KKPTSTKKPKTMPVRKPKTTPTRKMTSTMPELNPTSRITAEAMLQTTTTPRNPQTPNSKLV 995
Db 1078 -----PAVGDGQOQMTFGAAQSVPP-----VTEAAVQ----- 1104
QY 996 EVNPKSEDAGAGGE 1010
Db 1105 EAAAAAAGAAAGERE 1119

RESULT 10
T25697
hypothetical protein F16F9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T25697
R:Fulton, B.
submitted to the EMBL Data Library, August 1996
A:Description: The sequence of C. elegans cosmid F16F9.
A:Reference number: Z20071
A:Accession: T25697
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-1229 <FUI>
A:Cross-references: EMBL:U67956; PIDN:AAB07691.1; GSPDB:GNO0028; CESP:F16F9.2
A:Experimental source: strain Bristol N2; clone F16F9
C:Genetics:
A:Gene: CESP:F16F9.2
A:Map position: X
A:introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 9.4%; Score 662; DB 2; Length 1229;
Best Local Similarity 27.1%; Pred. No. 2.2e-21;
Matches 314; Conservative 91; Mismatches 367; Indels 388; Gaps 54;

```
QY 73 RCFEFEREGRECDCAQCKYDKCCPDYSEFCAEVKDNKNRKTTPKPPVVDAGSG 132
Db 34 RDKIVKHIKNTCTCSC-----KCPVD-----APSNP----- 61

QY 133 LONGDFKVTTPDTSTQINKVSTSPKITTAKPINKPRPSLPNSDTSKETSIVNKETT-- 190
Db 62 -----FDVSTTISSINNDN-VDIGFS-----GDSNPTGSSWFQIEATVGGQIVKS 106

QY 191 -----VETKETTITNKOTS---TDGKEKTTSA-----KETQSIEKTS---A 225
Db 107 EHNIDSSVEVEKVTSTDSATNAPTIGKDSSTPEIITGIWINSKESVTDMSSTFRS 166

QY 226 KDLAPTSVLAKPTPKAETTTGKGPALTPKPTPTTPKEPASSTPKPTPTTIKSAPTTP 285
Db 167 TLLSPTELLTSP-----ETLVSTDSSTSEOTSPONTETIAS--PMEINITT--EATTS 218

QY 286 KEPAPTTTKS-----APTTPKEPAPTTTKEP---APTTPKE---PAPTTTKEPA 328
Db 219 VEPSVSTLASDETVTAIAESTTVIAEVSTTEPTTAESTTKKSTTKAPATTEPT 278

QY 329 PTTTKSAPTTPKEPAPTTPKPA---PTTP-----KEPAPTTPKPT---PTTPKEPAPTT 378
Db 279 PTTTTEEVTTTEAERTSTTSSTSEKPTPTPLDNKIAGATGKPEITHFPVGTTPNFDI 338

QY 379 KEPAPTTKEPAPTPAKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 433
Db 339 ATEPFFVAKSDKWLKTAETATQOTTETVTDGPEKETTKNVSIETPIITVPLVETST 398

QY 434 TKSAP-----TTTKEPAPT---TTKSAPTTPKEPSPTTKE---PAPT-----TP 472
Db 399 TSTASKESDGFHTTLKLVITADSDSTESATTVKPFNEETTTKSHVVVPKTKGKVTP 458

QY 473 K-----EPAPTTPKKPAP-----TTTKEPAPT---TTTKEPAPT---P 491
Db 459 KLELSFDEPTEIT-KAPHPGKLLKKTTHFVLSDNFARYSEAKENDDYNHLDNYHREK 517

QY 492 APTTPKEPAP-----TTTKEPAPTA--PKEPAPTT---PKETAPTTKKLPTTPPEKLA-- 540
Db 518 EPTTTESSSTTEVTTTEEPANGNPPTTENPTTTEOPTSTAESTTALPFTTEQVTVTE 577

QY 541 -PTTPEKAPPTPEELAPTTPEEPTPTTPEPAPTTPKAAAPNTPKEPAPTTTKEPAPTT 599
Db 578 EPTTAESTATQ---KPTTTOESVST---EKTSTTKA---STTEE--PTTIDEPTTTT 625

QY 600 PKEPAPTTKETAPTTPKGTAPTTLKEPAPTTTPKKAPKELAPTTTKEPTSTTSOKPAP- 658
Db 626 ---ESSTGATAPTELSTSEETTELKLTTE-----GSTTTEPTTTAIFAEAST 674

QY 659 ---TTPKGATPTTPKEPAPTTTPKEPAPTTPKGTAPTTLKEPAPTTT----- 701
Db 675 GIITDDETTSTSTTPEITSKE--IVTESAITQTSVSVVSSPTPQLPKRWKAIVNKF 732

QY 702 -----KKPAPKELAPTTKGP-----TSTTSOKPAPTTTPKETAPT 736
Db 733 KHNLEVKELKKLLKKESTSTTGSDSSETTVAENIDEVITTEKEKVQVPTTPEKTS 792

QY 737 TPKEPAPTTPKKPAPTTPE--TPPPTTSE--VSTPTTKEPTTIHKSDBSTPELSAETTP 793
Db 793 TTTQEEITTTTTTTEKTSKTTTEKPTTSSEATTETTSEPT-----TEST----- 838
```

```
QY 794 KALENSPKPGVPTTKTAAATKPMETTTAKDKTTERDLRTPPETTTA----- 840
Db 839 -----IVDTSSATTEESSTAETTTISAE---TSETTSESAAFTGSPEN 882

QY 841 -----APKMTKETATTTEKTSKITATTATTO-VTST--- 870
Db 883 TALOSSOKSEENESSAEKFCARRDFVFKKHTTVKPAETTSAAVASTTTTETITTEKS 942

QY 871 TTQDTTPEFKITLTKTLTAPKVT---TTKKTITITTEIMNKPEETAKPKDRATNSKATTPK 927
Db 943 TLTETTPTEATTNEVGTGPAFVGTAPVDETTINTLELLSK-----INNTQISQPK 992

QY 928 PQ-----KPTKAPKKPTSTKKPKTMRVVRKPKPTTTPPKMTSTTPELN 970
Db 993 PTDISKTDALSSLSLIGLSFTKAPMPTI-----HTTTDAFVATEASLN 1039

QY 971 -----PTSRIAEA 978
Db 1040 DGSDDKKIIDEAQPTDEIRRA 1059

RESULT 11
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favella, A.; Vaudin, M.
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: 221536
A:Accession: T34513
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GNO0021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/3504/1
```

Query Match 9.4%; Score 661; DB 2; Length 3507;
Best Local Similarity 23.0%; Pred. No. 6.6e-21;
Matches 278; Conservative 173; Mismatches 459; Indels 300; Gaps 53;

```
QY 27 LSSCAGRCGEGYSRD-ATC-----NCDYNCOHYMECCPDPKRVCTAE----- 67
Db 1781 MGSCGCKCMAGYTGATCICKIEEPKSDKTA-----CTDEWSRUCELEKKOCTVDEEV 1835

QY 68 -----LSCKRCFESFERGRCDC-----DAQ----- 89
Db 1836 PQCGACLPGHHPIINGTCOSLQISGLCAQKNCKNKAECIDIHPDSHCFCSPDGFIDGMI 1895

QY 90 CKKYDKC-----CPDYSEFC-----AEVKDNKNRKTTPKPPVVDVDEA 129
Db 1896 CDDVDECNAGCDDENTKCENTGTSFNCVCLGFKKVDKCVVDEKKQPNREKIEIDEE 1955

QY 130 GSGLDN-GDFKVTT---PDTSTTHKNKVSTSPKITTA-----KPINRPSLPPN--- 174
Db 1956 NSSSNSNGQEKPTTKGIVSSTSATSSSETTAEPHVTTISISTTTKDMTSSKS-PENVTM 2014

QY 175 SDTSKETSITVYKNTV-ETKETTTTKQSTDOCKETTSKETSISAKTQSTKSAKDLAP--- 230
Db 2015 SSESPEVSTSSSKSTTASETTVSTSPSSSEAPLASAPTTTEVITTESVKSTTPKEE 2074

QY 231 -TSKVLAKPTPKAETTKGAPALTPPKPTPTTPKEPASTTTPKEPTTIKS-APTTPKEP 288
Db 2075 SSSEITVKLSSKSPVTESSVKSPSTPS-TTSQSVSTVTPETSKSVLSSEAPVSTSP 2133

QY 289 APTTTSKAPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPK 348
```

Db	2134	TEVHT--SSETKPSLGSASTTGDNTSTTPTSTSLASVSKSTSAPECTGASVAPKVLGSLSPD	2192
Qy	349	KAPATPK--EPATTPKEPTPTTTPKEPATPK--APTTPKEPATPAKPPATTPKEPA	406
Db	2193	VSPSTKTTDATSTESVQASSETSGTSVKSTSESESHVTKLSTSSNPSSVPIVSPKT	2252
Qy	407	PTTPKEPATTTKPESTPTT--KEPATTTKSAPTTPKKEPATTTKSAP-----TTPKE	458
Db	2253	PTVPE-----STEQPTSTTPSQSLTPMNSNEVLTTSEPHVLSLSPDVOSQSTPPNN	2307
Qy	459	PSPTTTKEPATTPKEPATTPKKEPATTPKEPATTPKEPATTP-----503	
Db	2308	LSSESTVE---TPKTSSEVLSNSEPTTTPATTLSPDILLSTTNLSSOSTVSTEDRS	2363
Qy	504	---TKKPAPTAK-----EPATTPKETAPTTPKTLPTTPKEKLAPTPPEKP	547
Db	2364	EISSENSEKPT--SAPELVTSSVTHVASSDPDVPTES--SEPDDLTGSGSTENIPEASSKQT	2420
Qy	548	APTTPPELATTPPEETPTTPPEP-----APTTPKAAAPNTPKEPATTT-----591	
Db	2421	ISSTPTPDTTASEEPTKSTSMSPOLSTSNVLSSESTPSESS---KPVSSSTEGIS	2476
Qy	592	---PKPEAPT-----TPKEPAP-----TTPKETAPTTPKGTAPTTLKEPAT	630
Db	2477	VWTSTESKVPSTISSVLEEDLTKTTPSPILEETTTASSETPLEDSLTVSVRIHET	2536
Qy	631	TPKKPAPKELAPTTTKEPTSTTSDKPA-----PT-----TP--K	662
Db	2537	TSSENVKPESESTTTSESCKPSQBPAGILTVVVPVSSVSLTASEIEALTSNTPFKQ	2596
Qy	663	GTAP--TTPK-----EPATTPKEPATTPKGTAPTTLKEPATTPKPKAPKELAPTT	713
Db	2597	GRTPIITSPKSLVKSTSTPSTVSSPESBTKTKTVSTVSTTPTTEETTTSESLILTA	2656
Qy	714	TKGT--STTSDKAPITPKETAPTTPKEPATTPKPKAPTPPEPPTTSVSPPTTKE	772
Db	2657	PSKPTSTTSSSEAPATTPAKTSEKPSNVVSTSRKSTENVETSTSQSLESSTMSSTS-	2715
Qy	773	PTTIHKSPDESTP--ELSAEPTPKALE--NSPKEGVPVTKTPAAK-----PMTTAK	823
Db	2716	---SEPETNAPAVTVSSEASTTLEENSSSTSP-----TSSEASVKLSLFPESI	2766
Qy	824	DKTTERDLRTPTTTAAPKMTKETATTTTEKTESKITATTT-----QVT	868
Db	2767	VTVSSR---APABITMSSSEHREITVSESEPEIPLSTTVSPNVVTASSIPSEPI	2822
Qy	869	STTTQDHTP--FKIHT-----LKTTLTAPKVTTTKTITTTTMMKPETAPKDRATNSK	922
Db	2823	SSVTSSSTPVRVLITGTPDDLIVSVTPVSHGNRRQNITASSV---PSNSTSPIILPSES-	2878
Qy	923	ATTTPKPKPTKAPKKPTSTKKPKTMPVRKPK---TTPPRKMTSTMPELN-----PTSR	974
Db	2879	LFTQPDPPTTTTAKPATTSCKKCPSPSIOPAEFMFTTAP-----PPPSNGGYGEETNQ	2932
Qy	975	IAEAMLOTTT	984
Db	2933	EEOQVSTSTT	2942

RESULT 12

RESULT
T45025

T43025 mucin MUC5B tracheobronchial [imported] - human (fragment)

mucin MUC5B, tracheobronchial
C. species: *Homo sapiens* (man)

C;Species: Homo sapiens (man)
 Date: 21-Jan-2000 #sequence revision 21-Jan-2000 #text change 21-Jul-2000

C;Date: 21-Jan-2000

C;Accession: T45025

R; Desseyn, J.L.; Guyonnet-Dup

J. Biol. Chem. 272, 3168-3178, 1997

A:Title: Human mucin gene MUC5B, the 10.7

A: Reference number:

A;Reference Number: 442899; MOID:37100151
A:Accession: T45025

A;Accession: T45025

A; Residues: 1-3570 <DES>
A; Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
A; Experimental source: placenta
C; Genetics:
A; Gene: MUC5B

Query Match	9.2%	Score	646.5	DB	2	Length	3570
Best Local similarity	22.5%	Pred.	No. 2.8e-20				
Matches	359	Conservative	107	Mismatches	536	Indels	591
QY	64	CTAE-----LSCKGRCPESFERGECDDAQ-----CKKYP-----KCCPDYE	101				
DB	506	CRANPEVSIIDOVGOVLC-----SLETGLTKNEQDTGRNMCNFYNVRLCCDDY-	558				
QY	102	SFCAEVKDNKNKTK-----KPTPKPPVDEAGSLDNGDFKVI--TPDTSSTQ	149				
DB	559	SHCPSTLATSSTAPSTPGTWTILTKPTTATTASTGSTATASSQTAGTPHVSST-	617				
QY	150	HNKVSVPKTTAK-----PINRPSLPNSDTSKETSLTWNKETTVE-----KET	196				
DB	618	-----ATTPTVTSKAPFPSPGTAALPALRSTATTPTATFTALPSSSLGTWTRLSTQ	673				
QY	197	TTTNKQSTDCGKKTSAKETOSIEKSA-----KDLAPTSKVLAKEPTP-	240				
DB	674	TTPMATNSTATPSTPETVHTSTVLTTATTGATGCVATPSSTPGTAHTKVLTTTTC	733				
QY	241	-----KAET-----TTKCPALT-----TPKEPTPTP-----K	263				
DB	734	FTATPSSPGRAATLPWISTTTPTTGSTVTPSSIGTHTTPVLTTTTPVATGMSMA	793				
QY	264	EPASTTPKEPTPTIKSAPTTKEPAPTTTKAP--TTPKBPAPTT--KEPAPT--TP	316				
DB	794	TPSSSTGTPPSLTTATTATGSTNPSSPTCTPIPPVLTATTATPAATSSVTTP	853				
QY	317	KEPAPTTKBPAPTTKS-----ABTP-----KEPAPTTKPKP	350				
DB	854	SSALGTTHTPPVNTTATTGHRSLSPSHPTVCTAWTSATSGILGTHHTPESTGTSHTP	913				
QY	351	APTT-----PREPAPTTPEPTPTTKEPAPTTKEPAPTTKEPAPTAPEKAPTTKEPA	406				
DB	914	AATTGTTQHSPTALLSPHPSRSTTESPPGCTTTCGHT-----ATSRTTATATPSKTR	967				
QY	407	PTT--PKPE--APTPTK-----EPS-----	422				
DB	968	TSLLPSQTSAPITTVTVMGCEPQCAWSEWLDYSYMPGPGSGDFDYSNIRAAGAVC	1027				
QY	423	-----	422				
DB	1028	EOPGLECRAQAQGVFLRELQGVVCSLDGLVLCNRQVCKFKMCFNYELRVCNCG	1087				
QY	423	-----PTTPKEPAPTTTKSAPTTT--KEPAPTTKSAPTTPKPSPTTTKEPAPTTKEPAPT	478				
DB	1088	HCPSTPATSSATSPSTPGTWTILTELTATTTESTGSTATPTSLRTAP--PPKVLTT	1145				
QY	479	TPKKPAPTTKEPAPTTKEPAPTTKKPA-----PTAPK-EPAP-----	517				
DB	1146	TATTTETVTSK-----ATPSSPGTATAPALRSTATTPTATSVTPIPSSSLGTWTRLSTQ	1201				
QY	518	-TTPKETAPTPKLLTPTPE-----KLAPTTPEKAPTTPEELAPTTPEPTPTTPEE	570				
DB	1202	TTTTPTATMSTATPSTPTETAHTSVLTATATTGATGCVATPSSTPGTAHTTKVPVTTTT	1261				
QY	571	PAPTTPKAAAPNTPKEP-----ABTTPEKAPTT--TPKEPAPTTKETAPTTPKGTAPT-T	623				
DB	1262	GFTATP-SSSPGATLTPVWISTTTPTTTRGSTVTPSSIPGTHHTATVLTTTITTVATGS	1320				
QY	624	LKEPAPTTKPEAPKEAPLT-----TPKETSTTSUKAP-----TTPKGTAPT-	667				
DB	1321	MATPSSSTQTSPTPSLTTATTATGCSNTNFPSPGPPPPVLTATTATPAATSSVT	1380				
QY	668	TPKEPAPTTKEPAPT-----	683				

```
Db 1381 TPSSALGTHTPVPVNTATTHGRSLSPSSPHTVRTAWTSATSGLTGLTHITPEPSTGSH 1440
Qy 684 TPKGTAFTT-----LKEAP---TPPKPAKELAPTTTKGTSSTSDKPAFTTKETA 734
Db 1441 TPAATGTTQHSTPALSSPHSSRTTESPPS---GTTTGHITATSRRTATATPSKTR 1496
Qy 735 PTT--PKEP-----741
Db 1497 TSTLLPSSPTAPITVTVTMGCEPCAWSEWLDYSYMPGPGSGDDFTYSNIRAAGAVC 1556
Qy 742 -----741
Db 1557 EQPLGLECRAQAQPVPLRELQGVWCSLDGLVCRNREQVGFKMCFNVEIRVFCNYG 1616
Qy 742 -APTTKKKAPTTPEPPPTT---SVSTPTTKEPTTHKSPDESTPELSAEPKALE 797
Db 1617 HCPSTATSTPSTPGTTHLLEQTAAATTAATGTAIP-SSTP--GTAPPKVL 1673
Qy 798 NSPKPGVPTTKPAATKPEMTT-----AKDKTTERDLRTTPTT-----838
Db 1674 SQATTTATSSKATSSSPRTATTLPLVLTSTATKSTATSFPIPSSTLGTGTSQNRPPH 1733
Qy 839 -----TAAPKMKETA-TTEKTTESKITATTQVIST--TTQDTP-----FKITTLK 884
Db 1734 PMATMTIHPSSPTTHTSTVLTK-----ATTTRATSSMSTPSTPGTTHLTELTTAA 1789
Qy 885 TTTLAPKVTTKTITTTIMNKPEETA-----KPKDRAT--NSKATTPKQKP 931
Db 1790 TTTAALPHGTPSTPGTTHLLEPSTATVTVTGSTATASSTRATAGTLKVLUTSTATTP 1849
Qy 932 TRAPKPTSTKKPKT---MPVRKPKTTPPKMTSTMPELNPTSRIAEA--MLQTTTRP 986
Db 1850 TVISSRATPSSSPTATALPALRSTATTPTATSVTAI-----PSSSLGTAWTRLSTQTTT 1904
Qy 987 NQTPNSKLVENPKSEDAGGAEETHMLLRPHVFMPEVTPMDYLPVP---NOGIILN 1043
Db 1905 TATMSTATPSTPETVHTSTVLTTATTTGTGSVATPSSPTPGTAHTKVPTTTTGTAT 1964
Qy 1044 PMLSDFTNLCNGKPVGLTLRLNGTLVAFRGVFWMLSPSPSPARRI-----1092
Db 1965 PSSSPGALT--PPVWIST---TTPTRGS---TVIPSSIGTTHATVLTTLTTTVA 2015
Qy 1093 -----TEVWGIPIPIDVTRCNCBGT 1115
Db 2016 TGSMATPSSSTQSGTSPSLTTTATTATGST 2048

RESULT 13
I38346
elastic titin - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C:Accession: I38346
R:Label: S.; Koilmerer, B.
Science 270, 293-296, 1995
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330
A:Accession: I38346
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-7962 <RES>
C:Cross-references: EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017427
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q31

Query Match 9.0%; Score 635.5; DB 2; Length 7962;
Best Local Similarity 22.8%; Pred. No. 1.8e-19;
Matches 306; Conservative 112; Mismatches 475; Indels 451; Gaps 63;
Qy 76 ESPERG-----RBCDCAQCKKYDKCCPDYE--SFCAEVKDNKKNR-----TKKKTP 121
```

```
Db 5783 EDFEEQYVEREGYDEGEEMEEAYQREVIOVQKEVVEESHKRVPAKVPEKKAPPP 5842
Qy 122 ----KPPV---VDEASGLDNGDFKVT-----TPDSTSTQHNKVSYS-PKITT 161
Db 5843 PKVIKRPVIEIKETISRRMEEKVQVTKPEVSKKIVPOKPSRTPVQEEVIEVKVPAVT 5902
Qy 162 AKINRPSPUPPSDTSKETSJT--VNEKTTVETKETTTNKOTSDGKEKTTSAKETQ 218
Db 5903 KKWVISEEKMFASHTEEEVSVPVQKEIVTEEKIHVAVSKRVE-----PPPKVPE 5955
Qy 219 SIEKTSAKDLAPT--SKVLAKTPKAEETTKGPALETTTPKEPTTTPKPEAS-----TTPK 271
Db 5956 LPEKPAPEAVPVPKPKVEPPAKVPVPEPKP-VPEKKKPVVPKKEPAAPKVPVPEVK 6014
Qy 272 EPTPTTIKSAPTTPKEAP-----TTTKSAPTTPKEAPTTTKEPA-PTTPKE 318
Db 6015 KPVEEKIPVPVAKKKEAPPAKVPQGVVTEEKITIVTQREESPPAVPEIPKK 6070
Qy 319 PAPTTPKEAPTTKSAPTTPKPAFTTPKPAFTTP-KBPAPTTPKEPTT-----PKEPAPT 377
Db 6071 KVPEERK-PVPRKEEEVPPPKVPA--LPKKVPVEEKVAVVPVAKKAPPRAEVSKTV 6127
Qy 371 -----PKEAPT 377
Db 6128 VEEKRFVAEEKLSFAPVQREVTRHEVSAEESYSSEEEGVSVISVYREEREEREEAEV 6187
Qy 378 TK-----EPAPT-----PKPA-----PT 392
Db 6188 TEYVAMEEPEEYVVEEKHLIIISKVPAEPAEAVTERQEKIIVLPKIPAKIEEPPPAKVP 6247
Qy 393 APKK-----PAPTTPKE--PAPTTPKEP-APTTPKEPSPTTPKE---PAPTTPKSAP 438
Db 6248 APKKIYVEKKVPAVPKKEKVPVPKVPVEEKVPVEKKVPKVKIKMEEPILPAKYTEKHM 6307
Qy 439 TTTKEP--APTTPKSAP--TTTPKEPSPTTPKE-----PAPT-----471
Db 6308 ITQEEKVLVAVTKKAPPAKVPPEEKRAVPPEEKVLKPKREEEPPAKVTEPRKRVKE 6367
Qy 472 -----PK-EPAP-----477
Db 6368 EKVSIEAPKREPQPIKEVTIMEEKERAYTLLEEAVSVQREEEYBEYEDYKEFEYEPT 6427
Qy 478 -----TPPKKAPT--TPKEAPPTTPKPAFTTPKPAFTT 511
Db 6428 EYDQYEEYEEYEEYEEYITEPEKIPVKEPVEEPVPVTPKPAKPAKVLKAVP-E 6486
Qy 512 PKEPAPTTPKETAPTTPKLTTPTEK-----LAPTTPKEPAPTTPPEELAPTPEEPT 564
Db 6487 EKVPVP-IPKKLKPPPPK--VPEEPKVPKEKIHISITKREKEQVTP---AAKVPMPK 6540
Qy 565 PTPPEEPAPTTPKAAAP--NTPKEPAPTTPKEPA-----596
Db 6541 RVVAEEKVPVPKREVAPPPVVPVPEVPEKLEPEEFAFEVEEVTHVEEYLVVEEEYIHEEE 6600
Qy 597 PTPPKEPAPTTPKETAPTTPKGTAPTTLKBPAPTTPKPAKPAKELAPTTTKEPTSTT---652
Db 6601 FITEEVPVVPVK-VPEVPKVPPEE-KRPVPVPKKEAPPKVPVPEVPKPEEKVPLI 6658
Qy 653 --SDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPA--PTTPKPKAPKE 708
Db 6659 PKKEKP---PPAKVPVPEKVPV-BEKVPVPVPPKKVEAP-----PAKVPVPEKVPPEK 6708
Qy 709 L-----APTTPKGTSTTSOKPAPTTPKETAPT-TPKE---PAPTTPKPAFTTPEPTPT 760
Db 6709 KVPVPAPKKVEAPPKVPVPEKLLIPEEKPTVPKKEAPPKVPKREPPVPVPVVALPQ 6768
Qy 761 TSEVSPITTTKEPTTIHKSDESTPELSAETPKALENSKEPCVPTTKTTPAATKPEMTT 820
Db 6769 EEEVLFEIEIVPEEVLPEEEVLPPEEEVLPPEEEVPEEVEEVEEVEEVEEVEEVEEVEE 6827
Qy 821 TAKDKTTERDL--RTPTETTTAAPKMTKETATTTTEKTTESKITATTTQVSTTTTQDTP 878
```


Db 6828 EEEFVPEEVLPEVKVPVPAP-----VPEIKKKVTEKKVVIPKKEEAAPAKVPEVPK 6882
QY 879 KIITLTKITLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPKQKTKAPKKP 938
Db 6883 KVE--EKRIILPK-----EEEVLPEVTEEPPE-----EPISSEETPEEP 6920
QY 939 TSTKK-----PKTMPVRVK---PKT-TPTPRKMTSTMPELNPTSRIAEAMLOTTTRPNQT 989
Db 6921 PSTEEVEAVPRVPEVIKKAPEAPTVPKVEA--PPAKVSKKIPPEKVPVPVQKKEA 6978
QY 990 PMSKLVNPKSEDAGAGETPHMLRPHVFMPEVTPDMYLPVPNOGIINPMLSDE 1049
Db 6979 PPAKVPEV-----PKKVPKKVL--VPKKEAV-----7003
QY 1050 TNCNCKPVDGLTLRNGTLVAER 1073
Db 7004 -----PPAKGRVLEEKVSVAFR 7021

RESULT 14
T11622
extensin class 1 precursor - cowpea
C:Species: Vigna unguiculata (cowpea)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000
C:Accession: T11622; S54155
R:Arseñijevic-Maksimovic, I.; Broughton, W.J.; Krause, A.
Mol. Plant Microbe Interact. 10, 95-101, 1997
A:Title: Rhizobia modulate root-hair-specific expression of extensin genes.
A:Reference number: Z17301; MUID:97155574
A:Accession: T11622
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-489 <RS>
A:Cross-references: EMBL:X91836; NID:g1015936; PIDN:CAA62943.1; PID:g1015937
A:Experimental source: sub-species Red calcoona
R:Arseñijevic-Maksimovic, I.; Broughton, W.J.; Krause, A.
submitted to the EMBL Data Library, April 1995
A:Description: A class of root-hair specific extensins involved in rhizobium/legume inte
A:Reference number: S54155
A:Accession: S54155
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 326-489 <AR2>
A:Cross-references: EMBL:X86030; NID:g791149; PID:g791150
A:Gene: Ext263
C:Superfamily: hydroxyproline-rich glycoprotein
C:Keywords: glycoprotein; hydroxyproline
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-489/Product: extensin class 1 #status predicted <MAT>

Query Match 9.0%; Score 633; DB 2; Length 489;
Best Local Similarity 31.2%; Pred. No. 1.6e-20;
Matches 149; Conservative 33; Mismatches 252; Indels 44; Gaps 7;
QY 285 PKEPAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTKPEAP 344
Db 39 PKOTPPYYNAPYYKSPSP-----PSPSP-----PPVYVHKYPPYYKSPSP 83
QY 345 TTPKKAPATTKPEAPTTKPEPTTTPKPEAPTTKPEAPTTT---KEPAPTAKKPAPTT 401
Db 84 PSPSPPPYYKSPSPSPSPPPYYKSPSPS--PSPPPYYKSPSPSPSPSPPPYY 141
QY 402 PKEPATTKPEAPTTTKESPTTKPEAPTTTKSAPTTTKSAPTTTKSAPTTKPEPSP 461
Db 142 YKSPSPSPSPPPYYKSPSPSPSPPPYYKSPSP-----PPSPSPSPSP 137
QY 462 TTPKPEAPTTKPEAPTTKPKAPATTKPEAPTTKPEAPTTTKKPAATKAPKPAPTTK 521
Db 188 YYKSPSPSPSPPPYYKSPSPSPSPPPYYKSPSPSPSPPPYYKSPSPSPSPSPS 247
QY 522 ETAPTTPKLTPTTPEKLAATTPEKAPTTPEELAPTTPEEPTTPEEAPTTKAAAP 581

Db 248 PPPYYKSPSPSPSPPPYYKSPSPSPSPPPYYKSPSPSPSPPPYYKSPSP 307
QY 582 NTPEKAPATTKPEAPTTKPEAPTTKPEATTPKGTATTPKGTATTLKPEAPTTKPKAPKELA 641
Db 308 PSPSPPPYYKSPSPSPSPPPYYKSPSPSPSPPPYYKSPSPSPSPPPYYKSPSP 366
QY 642 PTTTKETSITSDKAPATTKGTATTPKPEAPTTKPEAPTTKGTATTLKPEAPTTT 701
Db 367 ----KSPSPSPSPPPYYKSPSPSPSPPPYYKSPSPSPSPPPYYKSPSPSP 422
QY 702 KKPAKELAPTTTKGTSTTSKAPATTTPKETAATTPKPEAPTTKPKAPATTTPETPP 759
Db 423 SP-----PPVYKSPSPSPSPPPYYKSPSPSPSPPPYYKSPSPSPSPSP 475

RESULT 15
C84672
hypothetical protein At2g27380 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84672
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: C84672
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-761 <STO>
A:Cross-references: GB:AE002093; NID:g5306260; PIDN:AAD41992.1; GSPDB:GN00139
A:Gene: At2g27380
A:Genetics:
A:Map position: 2

Query Match 8.9%; Score 632; DB 2; Length 761;
Best Local Similarity 31.1%; Pred. No. 2.7e-20;
Matches 217; Conservative 44; Mismatches 354; Indels 82; Gaps 33;
QY 238 PTPKAETTTKGPALTTT---KEPTT-TPKEPASTTPKEPTTTIKSAPTTKPEAPTTT 293
Db 69 PPPIQKPPTYSPPYPPPIQKPPTYSPPYPPPIQKPPTYSPPYPPPIQKPPTT 128
QY 294 KSAPTTKPEAPTTTKEPAP--TTPKEPATTTKPEAPTTTKSAPTTT---KEPAPT-- 345
Db 129 YSPPIYP-----PPPIQKPPTYSPPVKKPPV---QMPPTTYSPPPIKPPPVHKPPTTYS 181
QY 346 TPKKP--APTTPKEPATTTP---KEPTP--TTPKEPATTKEPATT-TPKEPAPTAPK 395
Db 182 PPVKPPVHKPPTTYSPPPIKPPPVHKPPTTYSPPPIKPPPVHKPPTTYSPPVKKPPVHK 241
QY 396 KPAPT-TPKEPATTTPKEPAP---TTTKESPTTPKEPATTTKSAPTTTKPEAPTTTKS 451
Db 242 PPTPIYSPPIKPPPVHKPPTTYSPPVKKPPVQTPPTPIYSPVKKPPVHKPPTTYSPP 301
QY 452 APPTKEPSTTTKEP-----APTTPKEPAPT--TPKKAPATTTPKEPATTTPKEAPTTTK 506
Db 302 VKSPVQKPPTTYSPPPIKPPPVQKPPTTYSPPPIKPPPVKKPPTTYSPPVKKP--PPVHKP 360
QY 507 PAF--TAPKEPAPT---TPKETAPTTP---KKLTPTTPEKLAATTTPKEPATT---PE 553
Db 361 PPIYSPVKKPPPVHKPPTTYSPPVKKPPPIQKPPTTYSPPPIKPPPLQKPPTTYSPPPI 420
QY 554 ELAPTTPEEPTTTPKEPAPT---TPKAAAP--NTPEKAPTTTPKEPATTTPKEPAPTTP 608
Db 421 KLPPVKKPPTTYSPPVKKPPPVHKPPTTYSPPVKKPPPVHKPPTTYSPPVKKPPVKKPPT 480
QY 609 KETAPTTPKGTATTLKPEAPT--TPKKAPKELAPTTTKETPTSTTSKAPATTTPKGTAP 666
Db 481 TYSPPVQV-----PPVOKPPTTYSPPVKKPPPIQKPPT---PTYSPPPIKPPPVKKPPTTY 532

Search completed: April 26, 2002, 16:33:15
Job time: 790 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:33:14 ; Search time 62.75 Seconds
(without alignments)
766.017 Million cell updates/sec

Title: AA8

Perfect score: 7064

Sequence: 1 MAWKTLPIYLLLSVFIQ.....ARAITRSGTSLSKVWVWVNC P 1311

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1180	16.7	5179	1	MUC2_HUMAN
2	950	13.4	1664	1	SLP1_CLOTH
3	798.5	11.3	1367	1	AMYH_YEAST
4	651	9.2	1255	1	MUC1_HUMAN
5	556.5	7.9	875	1	FPL1_MYTED
6	555.5	7.9	2700	1	ZAN_HUMAN
7	551	7.8	620	1	EXTN_TOBAC
8	533	7.5	1087	1	NFH_MOUSE
9	530.5	7.5	1162	1	TCNA_TRYCR
10	530	7.5	865	1	CPN_DROME
11	518.5	7.3	872	1	FPL1_MYTCO
12	509.5	7.2	662	1	MUC1_XENLA
13	503.5	7.1	1970	1	RPB1_HUMAN
14	498.5	7.1	1970	1	RPB1_MOUSE
15	497	7.0	831	1	NFH_RAT
16	493.5	7.0	467	1	RPB1_CRIGR
17	488.5	6.9	826	1	SSP2_PLAYO
18	488	6.9	1020	1	NFH_HUMAN
19	475.5	6.7	267	1	EXTN_MAIZE
20	472	6.7	2142	1	BAT2_HUMAN
21	471.5	6.7	5376	1	ZAN_MOUSE
22	468.5	6.6	634	1	HWPI_CANAL
23	467	6.6	817	1	VRP1_YEAST
24	455	6.4	1161	1	YJ9P_YEAST
25	454	6.4	797	1	VGIX_HSVB
26	448.5	6.3	670	1	VG50_HSVII
27	442.5	6.3	751	1	FPI_MVTGA
28	439.5	6.2	1083	1	T2D3_HUMAN
29	432.5	6.1	439	1	XP2_XENLA
30	432	6.1	3164	1	TEGU_HSV11
31	430.5	6.1	3421	1	TEGU_HSVB
32	426	6.0	2715	1	TRX2_HUMAN
33	424.5	6.0	1125	1	MAP4_MOUSE

34	424	6.0	307	1	SGS3_DROME	P02840 drosophila
35	424	6.0	1251	1	YQJ3_CAEEL	Q09550 caenorhabdi
36	422.5	6.0	1794	1	YAV1_SCHPO	Q10172 schizosacch
37	422	6.0	2774	1	MAPA_RAT	P34926 rattus norv
38	421	6.0	1229	1	NI21_HUMAN	Q9Y2N3 homo sapien
39	419.5	5.9	2476	1	ZAN_PIG	Q28983 sus scrofa
40	419	5.9	1185	1	DRPL_HUMAN	P54259 homo sapien
41	418.5	5.9	1411	1	TGOF_HUMAN	Q13428 homo sapien
42	411.5	5.8	907	1	VGP3_EBV	P03200 Epstein-Bar
43	410	5.8	1183	1	DRPL_RAT	P54258 rattus norv
44	408.5	5.8	2517	1	MCR2_HUMAN	Q9Y618 h nuclear r
45	407.5	5.8	3256	1	KI67_HUMAN	P46013 homo sapien

ALIGNMENTS

RESULT 1						
MUC2_HUMAN						
AC	Q02817; Q14878;	STANDARD;	PRT;	5179 AA.		
DT	01-JUN-1994 (Rel. 29, Created)					
DT	01-NOV-1997 (Rel. 35, Last sequence update)					
DT	20-AUG-2001 (Rel. 40, Last annotation update)					
DE	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).					
GN	MUC2 OR SMUC.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=Intestine;					
RX	MEDLINE=94132002; PubMed=8300571;					
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;					
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.					
RT	Identification of the amino terminus and overall sequence similarity					
RT	to prepro-von Willebrand factor.";					
RL	J. Biol. Chem. 269:2440-2446(1994).					
RN	[2]					
RP	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.					
RC	TISSUE=Colon;					
RX	MEDLINE=93016075; PubMed=1400449;					
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,					
RA	Kim Y.S.;					
RT	"The human MUC2 intestinal mucin has cysteine-rich subdomains located					
RT	both upstream and downstream of its central repetitive region.";					
RL	J. Biol. Chem. 267:21375-21383(1992).					
RN	[3]					
RP	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.					
RX	MEDLINE=91358717; PubMed=1885763;					
RA	Toribara N.W., Gum J.R. Jr., Cuihane P.J., Lagace R.E., Hicks J.W.,					
RA	Petersen G.M., Kim Y.S.;					
RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays					
RT	and polymorphism.";					
RL	J. Clin. Invest. 88:1005-1013(1991).					
CC	-1- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND					
CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A					
CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS					
CC	AGENTS AT MUCOSAL SURFACES.					
CC	-1- SUBUNIT: MULTIMERIC.					
CC	-1- SUBCELLULAR LOCATION: SECRETED.					
CC	-1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,					
CC	BRONCHUS, CERVIX AND GALL BLADDER.					
CC	-1- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR					
CC	INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).					
CC	-1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND					
CC	VARIES AMONG DIFFERENT ALLELES.					
CC	-1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT					
CC	OF SILKWORM HEMOCYTIN.					
CC	-1- SIMILARITY: CONTAINS 2 VWFC DOMAINS.					
CC	-1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).					

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>), or send an email to license@isb-sib.ch).

CC	DR	EMBL; L21998; AAB95295.1; -	DR	EMBL; M74027; AAA59875.1; -	DR	EMBL; M94131; AAA59163.1; -	DR	EMBL; M94132; AAA59164.1; -	DR	MIM; 158370; -	DR	InterPro; IPR000359; Cys_knot.	DR	InterPro; IPR000561; EGF-like.	DR	InterPro; IPR002400; GF_cysknot.	DR	InterPro; IPR001007; VWF.	DR	InterPro; IPR001846; Vwd.	DR	Pfam; PF00007; Cys_knot; 1.	DR	Pfam; PF00094; vwd; 4.	DR	PRINTS; PR00438; GFCYSKNOT.	DR	SMART; SM00214; VMC; 2.	DR	SMART; SM00011; VMC_def; 2.	DR	SMART; SM00216; VWD; 4.	DR	PROSITE; PS00022; EGF_1; UNKNOWN_1.	DR	PROSITE; PS01185; CTCK_1; 1.	DR	PROSITE; PS01225; CTCK_2; 1.	DR	PROSITE; PS01208; VWF; 2.	KW	Glycoprotein; Repeat; Signal.	FT	SIGNAL	1	20	POTENTIAL.	FT	CHAIN	21	5179	MUCIN 2.	FT	DOMAIN	1401	1747	APPROXIMATE REPEATS.	FT	REPEAT	1401	1416	1.	FT	REPEAT	1417	1432	2.	FT	REPEAT	1433	1448	3.	FT	REPEAT	1449	1464	4.	FT	REPEAT	1465	1471	5.	FT	REPEAT	1472	1478	6.	FT	REPEAT	1479	1494	7A.	FT	REPEAT	1495	1517	7B.	FT	REPEAT	1518	1533	8A.	FT	REPEAT	1534	1556	8B.	FT	REPEAT	1557	1572	9A.	FT	REPEAT	1573	1596	9B.	FT	REPEAT	1597	1612	10A.	FT	REPEAT	1613	1635	10B.	FT	REPEAT	1636	1651	11A.	FT	REPEAT	1652	1675	11B.	FT	REPEAT	1676	1683	12.	FT	REPEAT	1684	1699	13.	FT	REPEAT	1700	1715	14.	FT	REPEAT	1716	1731	15.	FT	REPEAT	1732	1747	16.	FT	REPEAT	1748	1763	17.	FT	DOMAIN	4815	4886	18.	FT	DOMAIN	4924	4991	19.	FT	DOMAIN	5075	5160	20.	FT	DOMAIN	5161	5222	21.	FT	DOMAIN	5223	5288	22.	FT	DOMAIN	5289	5354	23.	FT	DOMAIN	5355	5420	24.	FT	DOMAIN	5421	5486	25.	FT	DOMAIN	5487	5552	26.	FT	DOMAIN	5553	5618	27.	FT	DOMAIN	5619	5684	28.	FT	DOMAIN	5685	5750	29.	FT	DOMAIN	5751	5816	30.	FT	DOMAIN	5817	5882	31.	FT	DOMAIN	5883	5948	32.	FT	DOMAIN	5949	6014	33.	FT	DOMAIN	6015	6080	34.	FT	DOMAIN	6081	6146	35.	FT	DOMAIN	6147	6212	36.	FT	DOMAIN	6213	6278	37.	FT	DOMAIN	6279	6344	38.	FT	DOMAIN	6345	6410	39.	FT	DOMAIN	6411	6476	40.	FT	DOMAIN	6477	6542	41.	FT	DOMAIN	6543	6608	42.	FT	DOMAIN	6609	6674	43.	FT	DOMAIN	6675	6740	44.	FT	DOMAIN	6741	6806	45.	FT	DOMAIN	6807	6872	46.	FT	DOMAIN	6873	6938	47.	FT	DOMAIN	6939	7004	48.	FT	DOMAIN	7005	7070	49.	FT	DOMAIN	7071	7136	50.	FT	DOMAIN	7137	7202	51.	FT	DOMAIN	7203	7268	52.	FT	DOMAIN	7269	7334	53.	FT	DOMAIN	7335	7400	54.	FT	DOMAIN	7401	7466	55.	FT	DOMAIN	7467	7532	56.	FT	DOMAIN	7533	7598	57.	FT	DOMAIN	7599	7664	58.	FT	DOMAIN	7665	7730	59.	FT	DOMAIN	7731	7796	60.	FT	DOMAIN	7797	7862	61.	FT	DOMAIN	7863	7928	62.	FT	DOMAIN	7929	7994	63.	FT	DOMAIN	7995	8060	64.	FT	DOMAIN	8061	8126	65.	FT	DOMAIN	8127	8192	66.	FT	DOMAIN	8193	8258	67.	FT	DOMAIN	8259	8324	68.	FT	DOMAIN	8325	8390	69.	FT	DOMAIN	8391	8456	70.	FT	DOMAIN	8457	8522	71.	FT	DOMAIN	8523	8588	72.	FT	DOMAIN	8589	8654	73.	FT	DOMAIN	8655	8720	74.	FT	DOMAIN	8721	8786	75.	FT	DOMAIN	8787	8852	76.	FT	DOMAIN	8853	8918	77.	FT	DOMAIN	8919	8984	78.	FT	DOMAIN	8985	9050	79.	FT	DOMAIN	9051	9116	80.	FT	DOMAIN	9117	9182	81.	FT	DOMAIN	9183	9248	82.	FT	DOMAIN	9249	9314	83.	FT	DOMAIN	9315	9380	84.	FT	DOMAIN	9381	9446	85.	FT	DOMAIN	9447	9512	86.	FT	DOMAIN	9513</
----	----	-----------------------------	----	-----------------------------	----	-----------------------------	----	-----------------------------	----	----------------	----	--------------------------------	----	--------------------------------	----	----------------------------------	----	---------------------------	----	---------------------------	----	-----------------------------	----	------------------------	----	-----------------------------	----	-------------------------	----	-----------------------------	----	-------------------------	----	-------------------------------------	----	------------------------------	----	------------------------------	----	---------------------------	----	-------------------------------	----	--------	---	----	------------	----	-------	----	------	----------	----	--------	------	------	----------------------	----	--------	------	------	----	----	--------	------	------	----	----	--------	------	------	----	----	--------	------	------	----	----	--------	------	------	----	----	--------	------	------	----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	------	----	--------	------	------	------	----	--------	------	------	------	----	--------	------	------	------	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	------	------	-----	----	--------	--------

[illegible]

Query Match	16.7%;	Score 1180;	DB 1;	Length 5179;
Best Local Similarity	30.4%;	Pred. No. 4e-41;		
Matches	389;	Conservative	68;	Mismatches 512; Indels 310; Gaps
QY	4	KTLPIYLLLLSVFVTQQVSSDLSGC-----AGRCG-----EGYSRDATCNDYNCQHWMEC	56	
DB	1169	KDRPIY-----EEDLKKCVTADKCGYVEDTHYPGASVPTTECKSCV-C	1213	
QY	57	CPDFKRVCTAELCKGRCFESFERG-----RECDCDAOCKKYDKC-----PDYES	102	
DB	1214	TNSQVVCRRPE-----EGKILNOTDGAFCYWEICGPNGVTERKHFNICSITTRPSTLITFTT	1270	
QY	103	FCAEVKDNKNRKKKPTPKPPV-----DEAGSGLDGDFK-----	139	
DB	1271	ITLPTPTSTFTTTTTPPSSIVLSTTPKLCCLWSWDINEDHPSSGSDGDRPFDGVC	1330	
QY	140	-----VTPDPTSTTOH-KNVSTSPK-----	158	
DB	1331	GAPEDIECRSVKDPHLSEHQGVQCDVSVGFIKCNEDQFGNGFGLCYDYKIRVNCW	1390	
QY	139	-----IITAKPINRPSLSPNSDTSKETSITSVKNKETTIVETKETTITNKOTSDGKEKITS	213	
DB	-1391	PMDCIITTPSPPTTTPSPPTTTTTPPTTTPSPPT-----TTTTPTTTPPTTTPPTTT	1445	
QY	214	AKETQSIEKTSAKDLAPTSKVLAKPTAKPTAKTTKGALTTPKPEPT-----PTTPKEPAST	268	
DB	1446	TTP-----LPTT-----TPSPPISTTTTPPTTTPSPPTTTPSPPTTTPSPPTT	1489	
QY	269	TPKEPTPTTKSAP-TTPKEP-----APTPTKSAP-TTPKEAPTPTTKPEAPTTPKEP	319	
DB	1490	TTTTTPPTTTPSPPTTTPPTTTPASTTTLPPTTTPSPPTTTTTPPTTTPSPPTTTPPT	1549	
QY	320	-----APTPTTKPEAPTPTTKSAP--TTPKEAPTTPKKPAPTTPKEAPTTPKEPTPTT	370	
DB	1550	PTISITTLPTTTPSPPTTTTTPPTTTPSPPTTTPSPPTTITTTTTPPTTTPSPPTTTT	1609	
QY	371	PKEAPTTPKEAPTTPKEP-----APTAPKKPAPTTPKEAPTTPKEAPTTPKEPSP	423	
DB	1610	TTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSPPTT	1669	
QY	424	TTPKEAPTPTTKSAPTPTTKPEAPTPTTKSAP---TTPKEPSPTTTPKEAPTTPTPKEAPTPT	480	
DB	1670	TTTTTTPPTTTPSSIIITTPSPPTTTTTPPTTTPPTTTPSSPI-TTTTTTPSSITTTTPSPPTTM	1728	
QY	481	KKPAPTTPKEAPTTPKEAPTPTTKKAPTAPKAPAPTTP-----KETAPTTP-----	528	
DB	1729	TTPSPPTTTPSPPTTTTTPPTTTPPTTSSLLTTPLPSPITTPPTTTPSPPTTTPPTCPVCLCNW	1788	

QY 529 -----KKLTP----- 533
 Db 1789 TGWLDGKPNFHKPGGDTLIGOVCGPWAANISCRATMYDPVPIQGLQGTVVCDVSL 1848
 QY 534 -----TPKELAPTTPEKAPTTPEELAPTTPEE 562
 Db 1849 ICKNEQKGGVPMFACFLNYEINVOCCBVTQPTMTTITVENPTPTPTPTTTT 1908
 QY 563 PPTT-TPPEAPTTTPKAAANTPKAPTTKPEP-----APTTPKEPAPT-TPKETAPT 615
 Db 1909 PPTPTGTQPT 1968
 QY 616 PKGTAPTTLKEAPATTTPKAPKAPKELAPTTTKEPTSTISDKAP-TTPKGTAPTTKPEAP 674
 Db 1969 PTTTTTTPPT 2028
 QY 675 TTPKEPAPTTPKGTAPTTLKEAPATTTPKAPKAPKELAPTTTKEPTSTISDKAP-TTPKET 733
 Db 2029 TGTQPTT-TPT 2087
 QY 734 APTTPKEPAPTTPKAPATTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 793
 Db 2088 TTTVPTPTPTGTQPTT-TPT 2143
 QY 794 KALENSPKP-----GVPT-TKTPAAKPEMTTAKDKTTERDLR---TTP--ET 837
 Db 2144 TGTQPT 2203
 QY 838 TTAAPKMT---KETATTTTEKTESKINATTTQVSTTTQDPTPTPTPTPTPTPTPTPTPT 894
 Db 2204 TTVPTPTPTGTQPT 2258
 QY 895 T-KTITTEIMNKPEETAKPKDRATNSKATTPKQPKAP-KKPTSTKKPKTMDPRVK 952
 Db 2259 TGTQPT 2318
 QY 953 -PKTPTPKMTSTWPELNP---ISRIEAMLOTTR-PNQTPTNSKLVEVNPKSEDAGGAE 1008
 Db 2319 TTVPTPTPTGTQPT 2375
 QY 1009 GETPHMLLRPHVFEVPT 1027
 Db 2376 TQTP--TTTPTG 2375

RESULT 2
 SLPI_CLOTH STANDARD; PRT; 1664 AA.
 AC Q06852;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE CELL SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER PROTEIN 1).
 DE OLPB.
 GN Clostridium thermocellum.
 OS Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1515;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIB 10682;
 RX MEDLINE=93209931; PubMed=8458832;
 RA Fujino T., Beguin P., Aubert J.-P.;
 RT "Organization of a Clostridium thermocellum gene cluster encoding the
 RT cellulosomal scaffolding protein CipA and a protein possibly involved
 RT in attachment of the cellulosome to the cell surface.";
 RL J. Bacteriol. 175:1891-1899(1993).
 CC - SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.
 CC - SUBCELLULAR LOCATION: CELL WALL.
 CC - SIMILARITY: CONTAINS 4 S-LAYER HOMOLGY (SLH) DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC -----

DR EMBL; X67506; CAA47841.1; -
 DR InterPro: IPR001119; SLH.
 DR Pfam: PF00395; SLH; 3.
 DR PROSITE; PS01072; SLH_DOMAIN; 2.
 KW Cell wall; S-layer; Signal; Repeat.
 FT SIGNAL 1 28 POTENTIAL.
 FT CHAIN 29 1664 CELL SURFACE GLYCOPROTEIN 1.
 FT DOMAIN 36 763 4 X 156 AA APPROXIMATE REPEATS.
 FT REPEAT 36 191 1.
 FT REPEAT 207 363 2.
 FT REPEAT 409 565 3.
 FT REPEAT 607 763 4.
 FT DOMAIN 771 1377 APPROXIMATE TANDEM REPEATS OF
 FT T-P-S-D-E-P
 FT GLY/PRO/SER/THR-RICH.
 FT DOMAIN 1453 1494 SLH 1 (INCOMPLETE).
 FT DOMAIN 1495 1565 SLH 2.
 FT DOMAIN 1566 1625 SLH 3.
 FT DOMAIN 1626 1646 SLH 4 (INCOMPLETE).
 FT SEQUENCE 1664 AA; 178194 MW; 5F396695BA9FE74B CRC64;

Query Match 13.4%; Score 950; DB 1; Length 1664;
 Best Local Similarity 31.0%; Pred. No. 3.5e-32;
 Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;

QY 234 VLAKPT-KAETTTKGPALTTPKEPTTPKPEASTTPKEPTTIKSAPTTPKEPAPT 292
 Db 758 VVIQAPAKAASDEPIPTDTPSDETPS-----DETPS---DETPSDETPSD 804
 QY 293 TKSAPTTPKEPAPTTPKEPAPTTPKEPA-----PTTKPEPAPTTPKSAPTTPKEPAPT 348
 Db 805 EPTPTSEPTPEEPTDTPSDETPSDETPSDETPSDETPSDETPSDETPSEETPEEPTDTP 864
 QY 349 KPAPTTPKEPAPTTPKEPTTPKPEAPTTPKEPAPT-TPKEPAPTAPKAPPTTPKEAP 407
 Db 865 SDEPTPSDETPS--DETPS--DETP--SDEPTSEPTPEEPTDTPSDETPSDEPT 919
 QY 408 TTPKEPA--PTTKPEPAPT-TPKEPAPTTPKS--APTTPKEPAPT---TTKSAPTTPKEPS 460
 Db 920 SDEPTPSDETPSDEPTPEEPTDTPSDETPSDETPSDETPSDETPSDETPSDEPT 979
 QY 461 ---PTTKPEPAPT-TPKEPAPTTPKAPPTTPKEPAPTTPKEPAPTTPKAPPTAPKEP 515
 Db 980 PSDEPTPSDEPTPEEPTDTPSDETPSDETPSDETPSDETPSDETPSDETPSDEPT 1039
 QY 516 APT---TPKETAPTTPKALTTPKPEAPTTPKEPAPT---TPEELAPTTPKEPTPT-TP 568
 Db 1040 TPDSEPTPEE---TPEEPTDTPSDETPSDETPSDETPSDETPSDETPSDEPTSETP 1094
 QY 569 EEPAPTTPKAAANTPKAPPTTPKEPA--PTTPKEPAPT-TPKETAPTTPKGTAPTTLK 625
 Db 1095 EEPPTDTPSDETPSDETPSDETPSDETPSDETPSDETPSDEPTPEEPTDTPSDEPT 1154
 QY 626 EPAPTTPKAPKAPKELAPTTTKEPT-STTSOKPAPTTPKGTAPTTPKEPAPTTPKEPAPT 684
 Db 1155 EPTPS--DETPSD--EPTPSDEPTPEEPTDTPSDETPSDETPS--DETPS- 1208
 QY 685 PKGTAPTTLKEPAPT-TPKAPKAPKELAPTTTKEPTSTTSOKPAPT---TPKETAPTTPKE 740
 Db 1209 ---DETPSDEPTPEEPTPEEPI-----PTDTPSDETPSDETPSD--EPTPSDE 1253
 QY 741 PAPTTPKAPAPT-TPETPTPTTSEVSTPTTKEPTTIHKSPDESTPELSAETTPKALENS 799
 Db 1254 PTPS--DETPSDEPTPEEPTDTPSDEPT---PSDEPTD---SDEPTPSDEPTP 1305

```
QY 800 PKRGVPTTKTAAAT--KPEMTTAKDKTTERDLRTPTTTTAAAPKMTKETATTTTEKTE 857
DB 1306 SDEPTSDPTSETPEEIPDTPTSPSD---EPTSPDEPTSPDEPTSPS 1361
QY 858 SKTATTTQVSTTTQDTPTFKITLTKTTTAPKVVTTTKKTTTTEIMNKPEETAKPKDR 917
DB 1362 DEPTSPDEPTSETPEEPT---TTTPTPTPTTT---PTSG 1396
QY 918 ATNSKAT-----TPKQKPTKAP--KPTSTKKTKTTPRVKPKTTTTPRRM----- 962
DB 1397 SGGSGGGGGGGGGVTPSPPTPTTSKPTSPAP--TEIEPTSPDPVPAIGGEHRA 1453
QY 963 -----TSTMPELNPTSRIAEAMLQTTTRNPQTPNSKLVENVNKSEDAAGAGETPHM- 1014
DB 1454 YLRGYPDGSFRPERNITRAEAAVIF-----AKLL-----GAGESYGAOSASPYSD 1498
QY 1015 LLRPH-----VFME-----VTPDMDYLPVRVNOGI----- 1040
DB 1499 LADTHAAWAIFKATSGLFKGYDGTGFKPDQNTIRAEFATVVHLTKVKGQEIIMSKLA 1558
QY 1041 ---IINPMLSDETNICNG---KPVDCGLTTL-----RNGTLVAFRGHYFWMLSPTSPSPS 1087
DB 1559 TIDISNPKFDD---CVGHWAQAEFIEKLTSLGYSYGPDCGT-----FKPON 1600
QY 1088 PARRITEVWIGIPSIDVFTTRCNCEGKTFPFKD--SOYWRFTNDIKD 1132
DB 1601 YIKRSESV---ALINRALERGLNGAPKLFPPDVNYSYWAF-GDIMD 1642
RESULT 3
ID AMYH_YEAST STANDARD; PRT; 1367 AA.
AC P08640; P08068;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLUCOAMYLASE S1/S2 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-
GN STAI OR STAZ OR MAL5 OR YIN019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Bartell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churche C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; PubMed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RL STAL.";
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RN SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STA2 and SGA genes
from Saccharomyces cerevisiae.";
RL FEBS Lett. 239:179-184(1988).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-
CC GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS
CC WITH RELEASE OF BETA-D-GLUCOSE.
CC -1- SIMILARITY: TO S.POMBE SPBC215.13.
```

```
CC -1- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z38061; CAA86176.1; -
CC EMBL; M16164; AAA35014.1; -
CC EMBL; M16165; AAA35015.1; -
CC EMBL; X13857; CAA32069.1; -
CC PIR; B26877; B26877.
CC PIR; B26877; A26877.
CC PIR; S48478; S48478.
CC SGD; S0001458; MUC1.
CC Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1367
FT DOMAIN 210 1367
FT CARBOHYD 817 817
FT CARBOHYD 874 874
FT SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;
SQ
Query Match 11.3%; Score 798.5; DB 1; Length 1367;
Best Local Similarity 27.9%; Pred. No. 4.2e-26;
Matches 313; Conservative 108; Mismatches 523; Indels 177; Gaps 50;
QY 85 DCDAQCKYDK---CCPDYESCAEVK-DNKNRKKKPTPKPPVDE--AGSGLDNGDF 138
DB 174 DLSTGCNNYDNOGHSOTDFGFYWNIDCNGCGTKSSTTSSTSESSTTSSESST 233
QY 139 KVITPTSTTOHKNVSTSPKITAKPIN------RPSLPNSDTSKETSILTUNKET 190
DB 234 TSSTSESSTTSSTSESSTTSSTTAPATTTTSCKEKPTPTTSCKEKPTPPHDTT 293
QY 191 VETKETTNNKQTSQDGEKTTSAKTSQIEKTSKDLAPTSKVLAKPFPKAEIT--TKG 248
DB 294 PCTKKTTTTSK-TCT---KKTTPVPTPS-SSTESSAPV-----PTPSSSTTESSA 342
QY 249 PALITPKE-----PTTPKPEASTTPKEPTTTIKSAP-----TTKPEAPTTPKAPT 298
DB 343 PVTSSTESSAPVPTPSSSTTESAPVTSSTESSAPVTSSTESSAPVPTPSST 402
QY 299 TPKEPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKEP 358
DB 403 TESSAPVTS---TTESSAPVTS---STTESSAPVT---SSTESSAPVTSSTT 451
QY 359 APTTPKEPTP---TTPKEPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 411
DB 452 ESSAPVPTPSSSTTESAPVT---SSTESSAPV---VPTPSSSTTESAPVTSSTTES 507
QY 412 EPAPTTPKPSPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPT 467
DB 508 SSAPVPTPS-SSTESSAPAPVPTPSSSTTESAPVTSSTESSAPVPTPSSSTTES 566
QY 468 AP-----TTPKEPAPT-TPKKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPT 522
DB 567 TPVTSSTESSAPVPTPSSSTTESAPVT---PSSSTTESAPAPVPTPSSSTTES 622
QY 523 TAPTPPKLTPTTPKEKLAPTTPKPAPTTPKEAPLTTPKEAPLTTPKEAPLTTPKE 576
DB 623 SAPVT-----SSTESSAP-VPTPSSSTTESAPVPTPSSSTTESAPVPTPSSSTTE 677
QY 577 KAAAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPA 635
DB 678 SSSAPVT-----SSTESSAPVT---SSTESSAPVPTPSSSTTESAPVPTPSSST 729
QY 636 APKELAPTTTKEPTSTTSDDKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTP 691
```

```

Db 730 TESSAPVTPPS--SSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTE 788
QY 692 TLKEPAPTPPKPAPKELAPTTTKGPTSTSDKPAPTTPKETAPTTPKEPAP-----TTP 746
Db 789 SSSAPVPTPSSSTTESSAPVPTPSSSSNITSSAPSSSTPSSSTESSVPVPTPSSSTTE 848
QY 747 KKPAPTPPTPTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 801
Db 849 SSSAPVSSSTTESSAPVPTPSSSSNITSSAPSSI---PFSSTTSFSGT-TVTPSSK 904
QY 802 EFGVPTTKPAATKPEMITTAKDKTTERDLRTPTTAAAPKMKETATTTTEKTTESKIT 861
Db 905 YFGQTEISVSTTETTTIVPTKTTTSVTPPTSTTTTTCSTGTNSAGETSGCSPKTVT 964
QY 862 AT--TQVTTSTTODTTPFKITTLTKTTTLAPKVT---TKKTTITT-EIMKPKETAKPK 915
Db 965 TVPTTTTTTSVTSSTTTTTCSTGTNSAGETSGCSPKTTTIVPCSTSPSETA--- 1021
QY 916 DRATNSKATTPKQKP-----TKAPKKPTSTKKPKTMPRVRRPKTPT 958
Db 1022 -----SESTTSPPTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTV 1076
QY 959 PKMTSTMPELNPTSRIAEAMLOTTTRENQTPNSKLVENPKSEDAGAGETPHMLLRP 1018
Db 1077 P--SVTVTNTPTT-IITTVCSGT-----NSAGETSGCSPKTTVTTVPC 1120
QY 1019 HVFMEVTPDMYLPVRPNQGIINPMLSDETNICNGKPDVGLTLRNGTLVAFRGHYFW 1078
Db 1121 STGTGEYTTAATLVTTAVTTTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTV 1173
QY 1079 MSLPSPSPARRTEVWGIPSDITVFTRC-----NCEGKT 1115
Db 1174 TLAPSAVTPATN-----AVPTTITT--TECSAATNAAGET 1207.

RESULT 4
MUC1_HUMAN
ID MUC1_HUMAN STANDARD; PRT; 1255 AA.
AC P15941; P15942; P15931; P17626; Q14128; Q16442; Q16437; Q9Y4J2;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM) (PEMT)
DE (EPITALIN) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)
DE (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H23AG) (PEANUT-
DE REACTIVE URINARY MUCIN) (PUM) (BREAST CARCINOMA-ASSOCIATED ANTIGEN
DE DF3).
GN MUC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=90368716; PubMed=2394722;
RA Len M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL J. Biol. Chem. 265:15294-15299(1990).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=9020794; PubMed=2318825;
RA Lightenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT "Episialin, a carcinoma-associated mucin, is generated by a
RT polymorphic gene encoding splice variants with alternative amino
RT termini.";
RL J. Biol. Chem. 265:5573-5578(1990).
RN [3]
SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90368715; PubMed=1697589;
RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,

```

```

RA Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of human tumor-associated
RT polymorphic epithelial mucin.";
RL J. Biol. Chem. 265:15286-15293(1990).
RN [4]
SEQUENCE FROM N.A.
RX MEDLINE=9107524; PubMed=2268309;
RA Lancaster C.A., Peat N., Duhig T., Wilson D.,
RA Taylor-Papadimitriou J., Gendler S.J.;
RT "Structure and expression of the human polymorphic epithelial mucin
RT gene: an expressed VNTR unit.";
RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN [5]
SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276413; PubMed=2351132;
RA Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N.,
RA Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,
RA Keydar I.;
RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms.";
RL Eur. J. Biochem. 189:463-473(1990).
RN [6]
SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276414; PubMed=2112460;
RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
RA Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;
RT "A transcribed gene, containing a variable number of tandem repeats,
RT codes for a human epithelial tumor antigen. cDNA cloning, expression
RT of the transfected gene and over-expression in breast cancer
RT tissue.";
RL Eur. J. Biochem. 189:475-486(1990).
RN [7]
SEQUENCE FROM N.A.
RX MEDLINE=91033045; PubMed=1688329;
RA Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,
RA Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;
RT "Isolation and characterization of an expressed hypervariable gene
RT coding for a breast-cancer-associated antigen.";
RL Gene 93:313-318(1990).
RN [8]
PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=88330762; PubMed=3417635;
RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA Burchell J.;
RT "A highly immunogenic region of a human polymorphic epithelial mucin
RT expressed by carcinomas is made up of tandem repeats.";
RL J. Biol. Chem. 263:12820-12823(1988).
RN [9]
SEQUENCE OF 1-169 FROM N.A.
RX MEDLINE=90088473; PubMed=2597151;
RA Abe M., Siddiqui J., Kufe D.;
RT "Sequence analysis of the 5' region of the human DF3 breast
RT carcinoma-associated antigen gene.";
RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN [10]
SEQUENCE OF 1-109 FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=96183746; PubMed=8608966;
RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;
RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT transcriptase polymerase chain reaction of the MUC1 gene.";
RL Int. J. Cancer 66:55-59(1996).
RN [11]
SEQUENCE OF 1-89 FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=96181716; PubMed=8604237;
RA Yu C.-J., Yang P.-C., Shew J.-Y., Hong T.-M., Yang S.-C., Lee Y.-C.,
RA Lee L.-N., Luh K.-T., Wu C.-W.;
RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
RT tissues.";
RL Oncology 53:118-126(1996).

```

[12]

SEQUENCE OF 1-46 FROM N.A.
TISSUE-Breast carcinoma;
Buluwela L., Liu Q., Ludmani Y.A., Gomm J.J., Coombes R.C.;
RA Submitted (Oct-1992) to the EMBL/GenBank/DDJB databases.
CC -! FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CYTOSKELETON.
CC
CC -! SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM
IS ALSO PRODUCED.
CC
CC -! ALTERNATIVE PRODUCTS: VARIOUS VARIANT ARE PRODUCED BY ALTERNATIVE
SPLICING.
CC
CC -! TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL
TUMORS, SUCH AS BREAST CANCER.
CC
CC -! PTM: HIGHLY GLYCOSYLATED (N- AND O-LINKED CARBOHYDRATES AND SIALIC
ACID).
CC
CC -! POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT
VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE
MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.
CC
CC -! SIMILARITY: CONTAINS 1 SEA DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; J05582; AAA60019.1; -;	
EMBL; M32738; AAA35804.1; -;	
EMBL; M32739; AAA35806.1; -;	
EMBL; J05581; AAA59876.1; -;	
EMBL; M61170; AAB53150.1; -;	
EMBL; X52229; CAA36478.1; ALT SEQ.	
EMBL; X52228; CAA36477.1; ALT SEQ.	
EMBL; M35093; AAB59612.1; ALT_SEQ.	
EMBL; Z17324; CAA78972.1; -;	
EMBL; Z17325; CAA78973.1; -;	
EMBL; M31823; AAA35757.1; -;	
EMBL; S81781; AAD14376.1; ALT_INIT.	
EMBL; S81736; AAD14369.1; ALT_INIT.	
EMBL; M21868; AAA59874.1; ALT_SEQ.	
PIR; A35175; A35175.	
PIR; B35175; B35175.	
PIR; S10218; S10218.	
GlycoSuiteDB; P15941; -;	
MIM; 158340; -;	
MIM; 113720; -;	
InterPro; IPRO00082; SEA.	
Pfam; PF01390; SEA; 1.	
SMART; SM00200; SEA; 1.	
PROSITE; PS0024; SEA; 1.	
KW Glycoprotein; Signal; cytoskeleton; Actin-binding; Transmembrane;	
Repeat; Alternative splicing.	
FT SIGNAL 1 23	POTENTIAL.
FT CHAIN 24 1255	MUCIN 1.
FT DOMAIN 24 1162	EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1163 1186	POTENTIAL.
FT DOMAIN 1187 1255	CYTOLASMIC (POTENTIAL).
FT DOMAIN 81 960	44 x 20 AA TANDEM REPEATS.
FT DOMAIN 1034 1151	SEA.
FT CARBOHYD 957 957	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 975 975	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1029 1029	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1055 1055	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1133 1133	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 19 19	T -> TATTAPKAT (IN ISOFORM B).
FT VARSPIC 20 22	MISSING (IN ISOFORM C).
FT VARSPIC 20 31	MISSING (IN ISOFORM D).
FT VARSPIC 126 905	MISSING (IN POLYMORPHIC EPITHELIAL ISOFORM).
FT VARSPIC 1077 1087	FLOIYKGQGLF -> VSIGIKSPMLP (IN SECRETED ISOFORM).

FT	VARSPLIC	1088	1255	MISSING (IN SECRETED ISOFORM).
FT	CONFLICT	2	2	T -> A (IN REF. 11).
FT	CONFLICT	134	134	P -> Q (IN REF. 9).
FT	CONFLICT	154	154	P -> Q (IN REF. 9).
FT	CONFLICT	1021	1021	S -> T (IN REF. 3).
FT	CONFLICT	1251	1251	A -> T (IN REF. 3).
FT	SEQUENCE	1255 AA; 122072 MW; 5E28DFC4DE7D9A82 CRC64;		
Query Match 9.2%; Score 651; DB 1; Length 1255;				
Best Local Similarity 27.8%; Pred. No. 3.9e-20;				
Matches 295; Conservative				
Qy	181	TSLVNKEITVETETITNNKQITDCKEKTSSAKETOSIEKTSAKD-LAPTISKVLAKPT	239	
Db	16	TVLVIV-----VTGSGHASSTPGGEKETSATQSSVPSSTKNAVSMTSSVLSSH	65	
Qy	240	P-KAETTTKGP--ALTTTPKEP-----TPTTPKEPASTTKPEPTPTTI---K	279	
Db	66	PCSGSSITQGQDVTLPATEPASGAATWGQDVTSVPVTRPALGSTTTPAHDVTSAPDNK	125	
Qy	280	SAPITPKEPAPITTKSAPITPKEPAPITTKP-----APTTPKEPAPITTKPEAPITTK	333	
Db	126	PAPGSTAPPAGHVT-SAPDT--RPAGSTAPPAGHVT-SAPDT--RPAGSTAPPAGHVT-	179	
Qy	334	SAPITPKEPAPITTP-----KKPAPITPKP-----APTTPKEPTPTTP-----	371	
Db	180	SAPDTRPAPGSTAPPAGHVT-SAPDTRPAPGSTAPPAGHVT-SAPDTRPAPGSTAPPAGHVT	239	
Qy	372	-----KEPAPITKEP-----APTTPKEPAPITAPK-----KPAPITPKPEAPITPK	412	
Db	240	SAPDTRPAPGSTAPPAGHVT-SAPDTRPAPGSTAPPAGHVT-SAPDTRPAPGSTAPPAGHVT	299	
Qy	413	PAPITTTKEPSPITPKPEAPIT--TKSAPITTKPEAPITTKSAPITPKPSPITTKP--	467	
Db	300	SAPDTRPAPGSTAPPAGHVT-SAPDTRPAPGSTAPPAGHVT-SAPDT--RPAGSTAPP	356	
Qy	468	----APTTPKEPAPITPK-----KPAPITPKPEAPITPK-----EPAPITTKKAPITAPKPE	515	
Db	357	GVTSPADTRPAPGSTAPPAGHVT-SAPDTRPAPGSTAPPAGHVT-SAPDTRPAPGSTAPP	416	
Qy	516	APTTPKETAPITPKLITTPTEKLAPITPEKPAITPEELAPITPEPTPTTPEEPAPIT	575	
Db	417	GVTSPADTRPAPGSTAPPAGHVT-SAPDTRPAPGSTAP-----PAHGVTSAPDTRPAPG	472	
Qy	576	PXA-----AAPNTPKPEAPITPK-----EPAPITPKPEAPITPKETAPITPKGTAPITLKEP	627	
Db	473	PPAHGVTSAPDTRPAPGSTAPPAGHVT-SAPDTRPAPGSTAP-----PAHGVTSAPDTRPAPG	528	
Qy	628	APTTP-----KKPAKELAPITTKPEPTSTSDKPAP--TTPKGITAPITPKPEAPIT	675	
Db	529	GSTAPPAGHVT-SAPDTRPAPGSTAP--PAHGVTSAPDTRPAPGSTAPPAGHVT-S-----APD	583	
Qy	676	TPKEAPITTPKG-----TAPITLKEAPITTP-----KKPAKELAPITTKGPTSTT	721	
Db	584	TRPAGSTAPPAGHVT-SAPDTRPAPGSTAPPAGHVT-SAPDTRPAPGSTAP--PAHGVTSAP	642	
Qy	722	SDKPAP--TTPK-----ETAPITPKPEAPITTP-----KKPAPITTPPTPTTSEVS	765	
Db	643	DTRPAPGSTAPPAGHVT-SAPDTRPAPGSTAPPAGHVT-SAPDTRPAPG--TAPPAGHVT	700	
Qy	766	TPPTTKPEPTI-----IHKSPDESITPELSAETPKALENSKPEGPVTTTKTPAATKPEM	818	
Db	701	APDTRPAPGSTAPPAGHVT-SAPDTRPAPGSTAPPAGHVT-SAP-----DTRPAPGSTAP--	753	
Qy	819	TTTAKDKTTERLDRITPETT-----TAAPKMTKETATTTEKTESKIIATTTQVITSTT	871	
Db	754	--PAHGVTSAPDTRPAPGSTAPPAGHVT-SAP-----DTRPAPGSTAPPAGHVT-SAPDTRP	807	
Qy	872	TQDTP--FKLITLTKTTLAPKVTTTKTITTT--EIMNKEETAKPKDRATNSKATTPK	927	
Db	808	PGSTAPPAGHVT-SAPDTRPAPGSTAPPAGHVT-SAPDTRPAPGSTAPPAGHVT-SAPDTRP	867	

QY 928 POKPTKAPKKTSTKPKTMRVRKKTPTPTPKMTSTMPBLNPTSRIAEAMLOTTRPN 987
 Db 868 P--GSTAPPAGVTSAPDTRP---APGSTAPPAGVTSAPDTRP-----PG 909
 QY 988 QT-PNSKLVEYNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPML 1046
 Db 910 STAPPAGVTSAPDTRPAGSTAPPAAH-----GVTSAPDNRRLGASTA---PPV 955
 QY 1047 SDETICNGKPVGDLTTLRNLTVAFRGHYFWMLSPPSPS 1087
 Db 956 HNVTSAGSAGSASTLVHNGTSARATTPASKSTPPFSIPS 996

RESULT 5

FPL_MYTBD
 ID FPL_MYTBD STANDARD; PRT; 875 AA.
 AC Q25460;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ADHESIVE PLAQUE MATRIX PROTEIN (POLYPHENOLIC ADHESIVE PROTEIN) (FOOT
 DE PROTEIN 1) (MEF1) (FRAGMENT).
 GN FPL
 OS Mytilus edulis (Blue mussel).
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
 OC Mytiloidea; Mytilidae; Mytilus.
 OX NCBI_TaxID:6550;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91025829; PubMed-1367451;
 RA Filpula D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.;
 RT "Structural and functional repetition in a marine mussel adhesive
 RT protein.";
 RL Biotechnol. Prog. 6:171-177(1990).
 RN [2]
 RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.
 RX MEDLINE-83135732; PubMed-6298211;
 RA Waite J.H.;
 RT "Evidence for a repeating 3,4-dihydroxyphenylalanine- and
 RT hydroxyproline-containing decapeptide in the adhesive protein of the
 RT mussel, Mytilus edulis L.";
 RL J. Biol. Chem. 258:2911-2915(1983).
 CC -I- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
 CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
 CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
 CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
 CC -I- SUBCELLULAR LOCATION: SECRETED.
 CC -I- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
 CC -I- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
 CC -I- PTM: THE DECAPEPTIDE A-K-P-S-Y-P-P-Y-K IS POST-TRANSLATIONALLY
 CC MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE
 CC HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE
 CC (DOPA) DERIVED FROM TYROSINE.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 or send an email to license@isb-sib.ch).

EMBL; X54422; CAA38294.1;
 DR InterPro; IPR002964; Adhesive_plaq.
 DR InterPro; IPR002965; P_rich_extensn.
 DR PRINTS; PR01216; ADHESIVEI..
 DR PRINTS; PR01217; PRICHEXTENS..
 KW Repeat; Hydroxylation.
 FT NON_TER 1
 FT DOMAIN 67 870 TANDEM REPEATS OF Y-K-[PA]-K-[LP]-[ST]-Y-
 P-P-[ST]
 SEQUENCE 875 AA; 100412 MW; 6EA85312748CAACE CRC64;

Query Match 7.9%; Score 556.5; DB 1; Length 875;
 Best Local Similarity 28.7%; Pred. NO. 2e-16;
 Matches 272; Conservative 106; Mismatches 375; Indels 195; Gaps 56;
 QY 148 TQNKVSTSKITAKPINP-----RP--SLPPNSDTSKETSILTINKETIVET 193
 Db 1 TKHEPVYKPKTSYSAKYKPPYQPLKKKVDYRPTKSYPPYGV-SKTNVLP LAKKLSSVYK 59
 QY 194 KEITTTNKOTISDVG--KEKT---SAKETQSTKTSKAKOLAPTSKVLAKPTKAEITTKG 248
 Db 60 PIKTTYNAKINPPVYKPKMTYPTTYKPKSYPPYTKSKPTTYKPKITYPPTYKAKPSY-- 117
 QY 249 PALTTPEKPTPTPKPE---PASTTPKEPTPTTIKSAPTP-----KEPAPTTTSAPPT 299
 Db 118 PSYKPKKTVYPTTKPKLTVPTTYKPKPSYPPYTKPKPSYKTKTKTTPSYKAKPSY 177
 QY 300 PKPEAPTTTKEPA--PTTTPKEPA-----PTTTKEPA--PTTTSAPTTPK--EPAPTT 346
 Db 178 P-----PTYKAKPSYPTTYKAKPSYPPYTKAKPTYKAKPTYPSTYKAKPSYPPTYKAKPTY 233
 QY 347 PKKPA--PTTTPKEPA--PTTTPKEPT--PTTTPKEP--APTTPKEAPTTTPKEPA--PTAPK 395
 Db 234 KAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPTYKAKPTYKAKPTYKAKPSYPPTYKA 293
 QY 396 KPA--PTTTPKEPA--PTTTPKEPA--PTTTPKEPS--PTTTPKEPA--PTTTSAPTTTTPKEPA- 445
 Db 294 KPSYPTTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYK 353
 QY 446 PTTTTSAPTTTPKEPSPTTTPKEAPTTTPKEAPTTTPKEAPTTTPKEPA--PTTTPKEPA--PTTTPKEPAP 501
 Db 354 PSTYKAKPSYPP--PTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSY 409
 QY 502 TTTKKPAPTAPKEAPTTTPKETAPTTPKKLTPTTPEKLTPTTPEKLTPTTPEKLTPTTPEKLTPTTPE 561
 Db 410 PPTYKAKPTYKAKP-----TYPST-YKAKPSYPP--PSYKAKPSYPPTYKAKPTYKA 457
 QY 562 EPTPTTPEAPTTTPKAAAPTTPKEAPTTTPKEAPTTTPKEA-----PAPTTTPKETAPT-- 615
 Db 458 KPT-----TPSTYKA-----KPSYPASAKPSYPPTYKSKSYSPSYKPKTYPPTYK 506
 QY 616 PKGTAPTTLKEAPAPTTPKKAPKELAPTTTKEPTSTTSDKAPTTTPKGATPTT-----PKE 671
 Db 507 PKLTVKPTYK-PKPSYPPSYKPKTYPPTYK-----PKISYPPTYKAKPSY 551
 QY 672 PAPTTPEAPTTTPKGT---APTTLKEPA--PTTTPKPA--PKELAPTTTTPG-PTSTSD 723
 Db 552 PATYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYKAKPTYPSTYKAKPSYPPTY 611
 QY 724 KPAPTTPKETAPTTPKPA--PTTTPKPA--PTT---PETTPPTTSEVSTPTTKE----- 772
 Db 612 KAKPSYPP--PTYKAKPSYPPTYKAPTTPSTYKAKPSYPPTYKPKISYPPTYKAKPSY 667
 QY 773 PTTTHKSDSTPELSAETPCKALENSPK-----EPGVPTT--KTPA-----ATKPEMTT 821
 Db 668 PPT-YKAKPSYPPTYKAKPTNPSTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPP 726
 QY 822 AKDKTTEDLRTTPETTAAPKMTKETATTTTEKTESKITATTQVTTSTTQDTPPFKIT 881
 Db 727 YKAPTTPKAPTTPSTYKAKP--TYKAKPTYPPTYKAK-----PSYPPTYKPKPSYPP 777
 QY 882 TLKTTTTLAPKVTTKKTTITTTIMNKPETAKPKDRATNSKATTPPKPQKTPKPK--P 938
 Db 778 TYKSKSIYSPSYKPKTYPPPT---YKPKLTYPTYK-----PKPSYPPSYKPKITYP 826
 QY 939 TSTKKPKTMRVRKPKTTPP-----RKMTSTMPNLNPTSR 974
 Db 827 STYKLKPSYPPTYKTSKTSYPPTYNKKISYSPSYKAKTSYPPAYKPTNR 874

RESULT 6

ZAN_HUMAN
 ID ZAN_HUMAN STANDARD; PRT; 2700 AA.

AC Q9Y493; 000218;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ZONADHESIN (FRAGMENT).
GN ZAN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE OF 1-2379 FROM N.A.
RX MEDLINE=99018118; PubMed=9799793;
RA Glocker G., Scherer S., Schattevoy R., Boright A., Weber J.,
RA Tsui L.C., Rosenthal A.;
RT "Large-scale sequencing of two regions in human chromosome 7q22:
RT analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci
RT reveals 17 genes.";
RL Genome Res. 8:1060-1073(1998).
[2]
RN SEQUENCE OF 2338-2700 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=97271566; PubMed=9126492;
RA Gao Z., Harumi T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RT zonadhesin gene (ZAN).";
RL Genomics 41:119-122(1997).
CC -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
CC -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOON OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC -!- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC -!- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.
CC -!- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF053356; AAC78790.1; -
DR EMBL; U83191; AAC51208.1; -
DR MIM; 602372; -
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000998; MAM.
DR InterPro; IPR002965; P_Rich_extensn.
DR InterPro; IPR002919; TIL.
DR InterPro; IPR003328; TILa.
DR InterPro; IPR001007; VWFD.
DR InterPro; IPR001846; Wvd.
DR Pfam; PF00629; MAM; 3.
DR Pfam; PF01826; TIL; 4.
DR Pfam; PF02345; TILa; 4.
DR Pfam; PF00094; wvd; 4.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00137; MAM; 2.
DR SMART; SM00214; VWC; 1.
DR SMART; SM00011; VWC_def; 3.
DR SMART; SM00216; WVD; 3.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS00740; MAM_1; 1.

DR PROSITE; PS0060; MAM_2; 4.
KW Glycoprotein; Transmembrane; Cell adhesion; Repeat.
FT NON_TER 1 109 MAM 1.
FT DOMAIN <1 136 MAM 2 (PARTIAL).
FT DOMAIN 112 326 MAM 3.
FT DOMAIN 161 326 MAM 4.
FT DOMAIN 322 446 MAM 4.
FT DOMAIN 483 951
FT 66 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
FT (MUCIN-LIKE DOMAIN).
FT VWFD 1 (PARTIAL).
FT DOMAIN 953 1065
FT DOMAIN 1066 1454
FT DOMAIN 1455 1861
FT DOMAIN 1862 2292
FT DOMAIN 2293 2684
FT DOMAIN ?
FT EGF-LIKE.
FT CARBOHYD 74 74
FT CARBOHYD 403 403
FT CARBOHYD 1023 1023
FT CARBOHYD 1099 1099
FT CARBOHYD 1618 1618
FT CARBOHYD 1737 1737
FT CARBOHYD 1832 1832
FT CARBOHYD 1878 1878
FT CARBOHYD 2136 2136
FT CARBOHYD 2505 2505
FT CARBOHYD 2374 2379
FT CONFLICT 2700 2700
FT NON_TER 2700 2700
SQ SEQUENCE 2700 AA: 293013 MW: 80E60C0B12277B1 CRC64;

Query Match 7.9%; Score 555.5; DB 1; Length 2700;
Best Local Similarity 32.6%; Pred. No. 5.5e-16;
Matches 199; Conservative 64; Mismatches 240; Indels 107; Gaps 32;

QY 164 PINPRSLPNSDTSKETSITVVKETVTETKNTKQTSIDGKEKTSKAKTSQSIKT 223
DB 445 PVKVLDELPPVSPVSS-----TGSEITGLTENPTISTK-----KPTVSIEKP 487
QY 224 SAKDLAPTSKVLAKPTPKAETTTKGPAITPKPTPTTPKPEASTTPKPT 278
DB 488 SVTTEKPT-----VPKEKPTIPTEKPTISTEK---PTIPSEKPNMPSEKPTIPSEKPTIL 539
QY 279 KSAPPTPKPEAPTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKSAPT 337
DB 540 TEKPTIPSE-KPTIPSEKPTISTEKPTVPTEE--PTPTTEETTYMEEPVPTKPSIPT 596
QY 338 TPKEAPPTPKKPAPTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTAKPK 397
DB 597 --EKPSIPTK---PTISMETIIISTEKPTICPEKPTIPTEK---PTIPTEKSTISPEK- 647
QY 398 APPTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 457
DB 648 -PTIPTE-KPTIPTEKPTISTEKPTISTEK-PTISPEKLTIPTEKTIPTI---EKPTIPT 701
QY 458 EPSPTTKPEAPTTKPEAPTTKPKPA-----PTTPKPEAPTTKPEAPTTKPKPAAPTAP 512
DB 702 EKPTISTEE--PTPTTEETIIISTEKPSIPMEKPTLPTEETTSVEETIIISTEKLTIPM-- 757
QY 513 KEAPPTPKETAPT-----TPKLTPTTPPEKLA-----PTTPKPEAPTTPEELAPTTPEE 562
DB 758 EKPTISTEKPTIPTEKPTISTEKLTIPTEKLTIPTEKLTIPTEKLTIPTEKLTIPTEKLTIPTE 814
QY 563 PTPTTPPEAPTTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 622
DB 815 KPTISPEKPTISTEK-----PTIPTE-KPTIPTE-----ETTISTEKLTIPTE 855
QY 623 TLKPEAPTTPKKPAKPELAPTTTKPTSTT-----SDKPAPTPKGTAPTTKPEAPTT 676
DB 856 --EKPTISPEKLTIPTEKPTISTEKPTIPTEKLTIPTEKLTIPTEKLTIPTEKLTIPTEKLTIP 911
QY 677 PKEAPTTPKGTAPTTLKPEAPTTKPKKPELAPTTTKGTSTTSKPKPTPKET--A 734
DB 912 PPHPSPTA-TGLAALVMSPHAPSPMTSV---ILGTTTTSRSSTGMSCP-PNARYESCAC 966


```

MEDLINE=91376547; PubMed=1896773;
RA Prioli R.P., Mejia J.S., Aji T., Aikawa M., Pereira M.E.A.;
RT "Trypanosoma cruzi: localization of neurominidase on the surface of
RL trypomastigotes.";
RL Trop. Med. Parasitol. 42:146-150(1991).
CC -1- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN
CC PARASITE INVASION OF CELLS.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC
CC LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACYLNEURAMINYL
CC RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACYLATED
CC NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,
CC GLYCOLIPIDS OR COLOMINIC ACID.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POSSIBLE).
CC -1- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPMASTIGOTES, MINIMUM
CC IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.
CC -1- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT
CC DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TNCA
CC PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.
CC -1- SIMILARITY: CONTAINS 3 BNR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: M61732; AAA30255.1; -
CC PIR: JH0557; JH0557.
CC HSP: P29768; IDIL.
CC DR InterPro: IPR002860; BNR.
CC DR Pfam: PF02012; BNR; 2.
CC KW Hydrolase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;
CC KW Phosphorylation.
CC FT DOMAIN 1 457 CYS-RICH.
CC FT REPEAT 23 34 BNR 1.
CC FT REPEAT 163 174 BNR 2.
CC FT REPEAT 209 220 BNR 3.
CC FT DOMAIN 458 588 FIBRONECTIN TYPE-III.
CC FT DOMAIN 589 1120 44 X 12 AA TANDEM REPEATS, LTR DOMAIN.
CC FT CARBOHYD 342 342 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 394 394 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SEQUENCE 1162 AA; 120032 MW; 07049221897C6A40 CRC64;
CC -----
Query Match 7.5%; Score 530.5; DB 1; Length 1162;
Best Local Similarity 29.4%; Pred. NO. 2.9e-15;
Matches 177; Conservative 100; Mismatches 212; Indels 113; Gaps 34;
QY 258 TPTPKK-PASTTPKKEPTTIIKSAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTP 316
DQ |||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 599 TPTPADSSAHSPTSPVDS---SAHSTSTPADSSAHSCTPTSPVDSSAHSCT---PSTP 651
QY 317 KE-PAPTTTKPEAPTTTKAPTTKPE-PAPTTKPKP-----APTTPKEPA---PTTP 363
DQ |||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DQ 652 ADSSAHSCTSTPVDSSAHSCTSTPVDSSAHSCTSTPVDSSAHSCTSTPADSSAHSCTPSTP 711
QY 364 KEPTP--TTPKEAPTTTKPEAPTTTKSAPTTTKPEPSTTTKEAPTTKPEPA---PT 478
DQ |||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DQ 765 HGTSTPADSSAHSCTPST---PADSSAHSCTPSTPVDSSAHSCT---PSTPVDSSAHSCTPS 817
QY 479 TP--KKPAPTTTKPEAPTTTKPEAPTTTKKPAAPKPEAPTTKPTAPTPKKL----TP 533
DQ |||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DQ 818 TPVDSSAHSCTSTPVDSSAHSCTPSTPVDSSAHSCTSTPADSSAHSCT--PSTPADSSAHSCTP 876

```

229	QY	APTSKVLAKP--TPK---AETITTKGPAULTTPKEPPTTPKEPASTTPKEPPTTTIKSAPT	284
8	Db	SPVSAPVAAPVTSPVAAPVQVVSAAVAP-----APAAFINVTPVAPPTLASVQPAT	61
285	QY	KPEAPPTTTKSAPTTP--KEPAPTTTKEPAPTTTPKEPAPTTTKEPAPT-----TTTSAPTTP	339
62	Db	VTIAPAPATAASVTPVASVAPVVAATP-----PAASPVSTPVAVAQIPVASAPVAP	116
340	QY	KEPAPTTPKKPAP-----ITPKEPAPTTKEPPTP-----TTPKEPAPTTKEPAPT	385
117	Db	PVAAATPTPVQIPVAAPVIATPPVAASAPTAAPVTPIVSPVIASTPPVPAANTTPVAAPV	176
386	QY	KPEPA--PTAPKKPAPTTTPKEPAPTTTPKEPAPTTTKE--PSPTTPKEPAP-----TTTSA	437
177	Db	AAVPAAPVVAPVLAAPAVAPVAVVAETPAPPPVAEIPVATIPCEVAPLIPEVSVATK	236
438	QY	PTTTPKEP---APTITKSAPTTPKEPSPTTTPKEPA--PTTPKEPAPTTPKKPAPTTKEPAP	493
237	Db	PLAASEPVPVAPPATITPVVPAASPHVSAPAVETAIVAVFSASTEPVVAANTLTAP	296
494	QY	TTP-----KEPAPTTTKKPAPKAPKEP---APTTPKET-----APTTPKKLTPT	534
297	Db	ETALAPVVAESQVAANTVVATPTTPEPPTIAPPVVAETEVASVAVAEITPVVPV	356
535	QY	TPEKL--APTTPKEPAPTTPEELAPTTPEEPTPTTPEEP---APTTPKAAAPNTKEPAPT	590
357	Db	AAESIPAPVVATTPVAT---LAVTDDVTASAVPELPPVIAPSPVSAVETPVDLAPP	413
591	QY	TPKEPAPTTTPKEPAPTTPKETATPTTP--KGAPTTTKE-----PAPTTPKKPAPKE	639
414	Db	V-----LPPVAAEPVPAVABETETAPASAPVTIAALDIPVAPVIAAPSDAPAPSA	469
640	QY	LAPTITKEPTTSDKPAPTTPKGTAPTTTP-----KEPAPTTTPKEP---APTTPKGTAPT	691
470	Db	AAPIVSTTPP---TTASVPEITAPPAAVPTEDIVSVLSERAAIETPVAPPVEVTVAAADV	527
692	QY	TLKEPA-----PTTPKKPAPKELAPTTT---KGPTSTTSDKPAPTT---PKETAPT	737
528	Db	APPEAAADLIIEPVPPAPIPDLLEQTSVPVAAEAESTSS--PIPETSLPPNEAVASP	585
738	QY	KPEAPAPTTPKKPAPTTPTETPPPTTSEVSTTTPPTTKEPTTIHKSPTD-----ESTPEL	787
586	Db	EVAVAPITAPETIP-----EPFSLATPTPEIPVAPVVIQEAADVAVPVTETSTSIPT	641
788	QY	SAEPTPKALENSPKEPGVPTTKTPAAKPEMTTAKTKTDERLRTTP--ETTTAAPKMT	845
642	Db	TVE--FPEAVAEKVLDPAI--TEAPVTTQEPDVANINDGAPATEI--TTPPAVEIVTAAAEVS	697
846	QY	KETATTTTEKTTESKITATTTQVTSF-----TTQDPTTPFKITTLKTTTLA--PKVITTTKKT	898

```

Query Match          7.3%; Score 518.5; DB 1; Length 872;
Best Local Similarity 29.2%; Pred. NO. 7e-15;
Matches 276; Conservative 80; Mismatches 407; Indels 181; Gaps 54;

Qy 125 VVDEAGSGSLGDFKVTTPDTST--TQH-----NKVSTSPKITTAKINRPSILPPNS 175
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 29 VYGSAYSGAGAYK-TLPGSHPGYSGKHVPVYKPMNKIPT-PYI---SKSPAPYKPKGY 84
QY 176 DTSKETSLVWNETTETETITNNKQSTDGKEKTSKQTSIEKTSKADLAPTSKVL 235
Db 85 YPTKRYQPTYGSKTNPPTTKYKIAKLSSYKAIKTTPYAKAKTSYPSYK-----HKIT 139
QY 236 AKETPKAETTTKGPALTTKPTPTTPKPEASTTTPKEPTTTIKSAPTTPKPEAPTTPK 294
Db 140 YPTPK-----PKITYP--PTYKQKPSYPPSYKPKITYPTTYK-----PKITYPTTYK 185
QY 295 SAPT-TPKPEAPT--TTKPEAPTTPKPEAPTTPKPEAPTTPKASAPTTPKPEAPT--PKK 349
Db 186 RPSYTPYKPKATTPYTPKPKITYP-----PTYKRPSTY-----PYKPKTTPYTPYKPKI 236
QY 350 PAPTTPKPEAPT--PKEPTPTT--PKEAPTTPKPEAPTTPKPEAPT-APKPEAPTTP 402
Db 237 SYPSYKPKASVSSYKSKTTPPTTKYKISYPTTYKPKPSY-----PTYKPKVTYPTTY 292
QY 403 KEAPTTPKPEAPTTPKPSPTTPKPEAPTTPKASPT-----TTKEAPTTPKASAPTTPK 457
Db 293 K-PKPSYPTTPYKPKITYPTT-YKPKPSYPTTYKQKPSYPTTYKSKSYPTSYKSKKTYPP 350
QY 458 EPSPTTTKEPA-PTTPKPEAPTTPKPKP-APT-TPK-----EPAPTTPKPEAPTTPK 507
Db 351 TYKPKITYPTTYKPKPSYPPSYKPKATYSPTYKPKITYPTTYKPKPSYPPSYKSKKTYPP 410
QY 508 APTAPKEAPTTPKETAP-----TTPKKLTPTTPEKLAPTTPKPEAPTTPPELAPTTPPE 562
Db 411 T-YKPKISYPTTYKPKASVSSYKSKTTPPTTYKPKISYPTTYKPKPSYPTTYKPKITYP 469
QY 563 PT-PTTPEAPTTPKKAAPTTPK-EPAPTTPKPEAPTTP-----KEPAPT--PKETAP 613
Db 470 PTYKPKPSYPTTYKPKITYPTTYKPKPSYPTTYKPKPSYPTTYKPKPSYPTTYKSKKTYPP 529
QY 614 TT--PKGTAPTTLKPEAPTTPKPKAPKELAPTTPKPTSTTSKPEAPT-TPKGTAPTTPK 670
Db 530 PTYKPKITYPTTYK-----PKPSYPPSYKPKITYPTTYKPKITYPTTYKPKASYPPTYK 583
QY 671 EPAPTTPKPEAPTTPKGTAPTTLKPEAPTTP-----KKPAPKELAPTTPKGTSTTSKPKA 726
Db 584 -----PKITYPTTYK-PKPSYPTTYKQKPSYPTTYKSKSYPTSYKSKKTY 628
QY 727 PTT--PKETAPTTPKPEAPTTPK--KPAPTTPET-----PPTTSEVSTPTTKEPTT 775
Db 629 PPTYKPKITYPTTYK-PKPSYPPSYKPKITYPTTYKPKPSYPTTYKPKPSYPPSYKPKK 687
QY 776 IHKSPDESTELSAEPTPKALENSKPEGVPPTTKTTPAATKPEMTTAKDKT-----TERDL 831
Db 688 Y---PPSYKPKKTYPTTYK-----PKISYPTTYK-----TKPSYPPSYKPKKTSYPTTYKPK 735
QY 832 RTTPTETTTAAPKMTKETATTTKTSKITATTTQVTTSTTQDTTPTKITTTLKTTTLAPK 891
Db 736 ISVPSTYKAKP-----SYPTTYKPKPSYASSYKPKITYPTTYKPKPSYASSYK-----PK 785
QY 892 VTITTKTITTTMKNPEETAKPKDRATNSKATTPKPKP---KPTKAPKPKP-TSTKPKP-T 946
Db 786 I-----RYPPTTKPKP-----SYASSYKPKIRYPPTTYKPKPSYASSYKPKKIT 827
QY 947 MPVRKPKKIT--PTPKMTSTMPKELNPTSRIAEAMLOTTTRPNQ 988
Db 828 YPTTYKPKISYPTTYKPKITYPTTYKPKISYPTTYKPKISYPTTYKPKISYPSQ 871

RESULT 12
MUC1_XENLA
ID MUC1_XENLA STANDARD; PRT: 662 AA.
AC Q05049;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 01-OCT-1994 (Rel. 30, Last annotation update)
DE INTEGRATORY MUCIN C.1 (FIM-C.1) (FRAGMENT).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=93077556; PubMed=1447205;
RA Hauser F., Hoffmann W.;
RT "P-domains as shuffled cysteine-rich modules in integumentary mucin
RT C.1 (FIM-C.1) from Xenopus laevis. Polydispersity and genetic
RT polymorphism.";
RL J. Biol. Chem. 267:24620-24624(1992).
CC -|- FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL
CC INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.
CC -|- SUBCELLULAR LOCATION: SECRETED.
CC -|- ALTERNATIVE PRODUCTS: A NUMBER OF DIFFERENT FORMS OF THE PROTEIN
CC MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -|- TISSUE SPECIFICITY: SKIN.
CC -|- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC -|- SIMILARITY: CONTAINS 6 P-TYPE (TREFOL) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L02115; AA074725.1; -
DR PIR: A45155; A45155.
DR HSP: P04002; INFA.
DR InterPro: IPR000519; P-trefol.
DR Pfam: PF00088; trefol; 6.
DR SMART: SM00018; P; 6.
DR PROSITE: PS00025; P-TREFOL; 6.
KW Repeat; Amphibian skin; Glycoprotein; Alternative splicing.
FT NON_TER 1
FT DOMAIN 81 144
FT 8 X 8 AA APPROXIMATE TANDEM REPEATS,
FT ALA/THR-RICH.
FT REPEAT 81 88
FT 1-1.
FT REPEAT 89 96
FT 1-2.
FT REPEAT 97 104
FT 1-3.
FT REPEAT 105 112
FT 1-4.
FT REPEAT 113 120
FT 1-5.
FT REPEAT 121 128
FT 1-6.
FT REPEAT 129 136
FT 1-7.
FT REPEAT 137 144
FT 1-8.
FT DOMAIN 161 202
FT P-TYPE 1.
FT DOMAIN 218 301
FT 8 X APPROXIMATE TANDEM REPEATS, THR-RICH.
FT REPEAT 218 224
FT 2-1.
FT REPEAT 225 239
FT 2-2.
FT REPEAT 240 249
FT 2-3.
FT REPEAT 250 259
FT 2-4.
FT REPEAT 260 275
FT 2-5.
FT REPEAT 276 287
FT 2-6.
FT REPEAT 288 294
FT 2-7.
FT REPEAT 295 301
FT 2-8.
FT DOMAIN 306 347
FT P-TYPE 2.
FT DOMAIN 353 394
FT P-TYPE 3.
FT DOMAIN 402 522
FT 12 X APPROXIMATE TANDEM REPEATS,
FT THR-RICH.
FT REPEAT 402 411
FT 3-1.
FT REPEAT 412 419
FT 3-2.
FT REPEAT 420 431
FT 3-3.
FT REPEAT 432 443
FT 3-4.
FT REPEAT 444 453
FT 3-5.
FT REPEAT 454 460
FT 3-6.
FT REPEAT 461 472
FT 3-7.
FT REPEAT 473 479
FT 3-8.
FT REPEAT 480 491
FT 3-9.
FT REPEAT 492 498
FT 3-10.
FT REPEAT 499 515
FT 3-11.

```



```

437 TTTT-----TTKAT-----TTTPTTTPTTTT---KAT--- 463
831 LRTTETTTAAKMTKEATTEKTESKIATTTQVTSITTTQDTPFKITTLAKTTTLP 890
464 -TTPTTTT---TTTUKAT-----TTPTTTTPTTTPTTTTATTTPTTTT 510
891 KVTITTKTITTTIIMNKDEETAKPDATNSKATPKQKPTKAPKPTSKKPK---T 946
511 TTTTTKATTTTTSGECKME----PSKRDCGYPGITESQCRSGKCCFDSSIPQTKWCFYS 566
947 MPVRVKPTTPTPR 960
567 LPQVADCKVAPSSR 580

RESULT 13
RPBL_HUMAN
ID ID RPBL_HUMAN STANDARD; PRT; 1970 AA.
AC P24928;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPBL).
GN POLR2A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
SEQUENCE FROM N.A.
MEDLINE=92178992; PubMed=1542581;
RA Winterth M., Acker J., Vicaire S., Vigneron M., Kedinger C.;
RT "Complete sequence of the human RNA polymerase II largest subunit.";
RL Nucleic Acids Res. 20:910-910(1992).
[2]
SEQUENCE FROM N.A.
MEDLINE=95347616; PubMed=7622068;
RA Mita K., Tsuji H., Morimyo M., Takahashi E., Nenoï M.,
RA Ichimura S., Yamauchi M., Hongo E., Hayashi A.;
RT "The human gene encoding the largest subunit of RNA polymerase II.";
RL Gene 159:285-286(1995).
-1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
SUBSTRATES.
-1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
RNA(N).
-1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
-1- SUBCELLULAR LOCATION: NUCLEAR.
-1- PTM: THE TANDEM 7 RESIDUE REPEATS CAN BE HIGHLY PHOSPHORYLATED.
THE PHOSPHORYLATION ACTIVATES POL2.
-1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
III FOR 5S AND TRNA GENES.
-1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).

EMBL; X63564; CAA45125.1; -
DR EMBL; X74874; CAA52862.1; -
DR EMBL; X74873; CAA52862.1; JOINED.
DR EMBL; X74872; CAA52862.1; JOINED.
DR EMBL; X74871; CAA52862.1; JOINED.
DR EMBL; X74870; CAA52862.1; JOINED.
DR EMBL; S21054; S21054.
DR MIM; 180660; -

```



```

DR InterPro: IPR000684; RNA_polII_repeat.
DR InterPro: IPR000722; RNA_pol_A.
DR InterPro: IPR002879; RNA_pol_A2.
DR Pfam: PF00623; RNA_pol_A_1.
DR Pfam: PF01854; RNA_pol_A2_1.
DR PROSITE: PS00115; RNA_POL_II_REPEAT; 43.
KW Transferrase: DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
FT ZN_FING 71 87 C2H2-TYPE (POTENTIAL).
FT DOMAIN 1590 1958 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
FT CONFLICT 1067 1067 W -> L (IN REF. 2).
FT CONFLICT 1449 1449 D -> Y (IN REF. 2).
SQ SEQUENCE 1970 AA; 217205 MW; 6876FC25692A657E CRC64;

Query Match 7.1%; Score 503.5; DB 1; Length 1970;
Best Local Similarity 33.1%; Pred. No. 5.5e-14;
Matches 181; Conservative 101; Mismatches 167; Indels 95; Gaps 47;

QY 257 PPTTPKEPASTPKE-PTPTTKSAPTTPKEPAPTPTTKSAPTTPKE-----PAPTTPKE 310
Db 1507 PPMGGISPAWTPWNGATPAYGAWSPVSGWMTPGAAGSFSAASDASGSPGSPAWS 1566

QY 311 PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 368
Db 1567 PTPGSGPGSPSSPIPSGCGAMSPSYSPSPA-YEPRSPGGYTPQSPSPSYSPSPA-SP 1624

QY 369 TTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 428
Db 1625 TSPSY-SPTSPSYSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY 1671

QY 429 PAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 488
Db 1672 -SPTSPSYSPSY-SPTSPSYSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY 1725

QY 489 KEPAATTPKEPAPTTPKPAATTPKEPAPTTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPA 548
Db 1726 SY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-T 1777

QY 549 PTTPEELAPTPPEPTTPEEPAPTTPKAAANTPKAPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 606
Db 1778 TSP-NYSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY 1828

QY 607 TPKEATTPKGAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKGA 665
Db 1829 SP-SYSPSPKPT-----TSPSYSPSPSYSPSYSPSYSPSYSPSYSPSYSPSYSPSYSPSYSPSYSPSY 1875

QY 666 PTPKEPAPTTPKEPAPTTPKGAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKGA 724
Db 1876 PTPSY-SPTTPKY-SPTSP-SPTSPSYSPSYSPSYSPSYSPSYSPSYSPSYSPSYSPSYSPSYSPSYSPSYSPSY 1922

QY 725 PAPTTPKETAATTPKEPAPTTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKE 782
Db 1923 YSPTSP-TYSPSPKSTYSPSPSY-SPTSP-----TYSUTSPA-SPDD 1965

QY 783 STPE 786
Db 1966 SDEE 1969

RESULT 14
ID RB1_MOUSE
AC P08775;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE, II LARGEST SUBUNIT (EC 2.7.7.6) (RB1).
GN POLR2A OR RPO2-1 OR RPL1215.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;

```

```

[1] SEQUENCE FROM N.A.
RP MEDLINE-87280135; PubMed-3038894;
RX Ahearn J.M. Jr., Bartolomei M.S., West M.L., Cisek L.J., Corden J.L.;
RA "Cloning and sequence analysis of the mouse genomic locus encoding
RT the largest subunit of RNA polymerase II.";
RL J. Biol. Chem. 262:10695-10705(1987).
RN [2]
RP SEQUENCE OF 1587-1970 FROM N.A.
RP MEDLINE-86068017; PubMed-2999785;
RX Corden J.L., Cadena D.L., Ahearn J.M. Jr., Dahmus M.E.;
RA "A unique structure at the carboxyl terminus of the largest subunit
RT of eukaryotic RNA polymerase II.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:7934-7938(1985).
RN [3]
RP REVISIONS, AND PRESENCE OF AN ADDITIONAL EXON.
RP MEDLINE-92178992; PubMed-1542581;
RX Winterlith M., Acker J., Vicaire S., Vigneron M., Keding C.;
RA "Complete sequence of the human RNA polymerase II largest subunit.";
RL Nucleic Acids Res. 20:910-910(1992).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
CC RNA(N).
CC -!- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- PTM: THE TANDEM 7 RESIDUE REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC THE PHOSPHORYLATION ACTIVATES POL2.
CC -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND TRNA GENES.
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M12130; AAA40071.1; -
DR EMBL; M14101; AAA40071.1; JOINED.
DR PIR; A28490; A28490.
DR MGD; MGI:98086; Rpo2-1.
DR InterPro; IPR000684; RNA_polII_repeat.
DR InterPro; IPR000722; RNA_pol_A.
DR InterPro; IPR002879; RNA_pol_A2.
DR Pfam; PF00623; RNA_pol_A_1.
DR Pfam; PF01854; RNA_pol_A2_1.
DR PROSITE; PS00115; RNA_POL_II_REPEAT; 42.
DR Transferrase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
DR DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
FT ZN_FING 71 87 C2H2-TYPE (POTENTIAL).
FT DOMAIN 1590 1958 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
FT CONFLICT 1498 1498 P -> R (IN REF. 1 AND 2).
FT CONFLICT 1499 1536 MISSING (IN REF. 1 AND 2).
SQ SEQUENCE 1970 AA; 217175 MW; 7D76F38FD92A657E CRC64;

```

```

Query Match 7.1%; Score 498.5; DB 1; Length 1970;
Best Local Similarity 33.1%; Pred. No. 8.7e-14;
Matches 180; Conservative 101; Mismatches 168; Indels 95; Gaps 47;

QY 257 PPTTPKEPASTPKE-PTPTTKSAPTTPKEPAPTTPKPAATTPKEPAPTTPKEPAPTTPKEPAPTTPKE 310
Db 1507 PPMGGISPAWTPWNGATPAYGAWSPVSGWMTPGAAGSFSAASDASGSPGSPAWS 1566

QY 311 PAPTTPKEPAPTTPKEPAPTTPKSAATTPKEPAPTTPKPAATTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 368
Db 1567 PTPGSGPGSPSSPIPSGCGAMSPSYSPSPA-YEPRSPGGYTPQSPSPSYSPSPA-SP 1624

```



```

Db 243 KEOGEEEEEDEGEEAATTS-----PPAEAAASP-----EKETKSPVKEEAKSPAEAKS 293
QY PTTTPKEPA-STTPKEPTPTTIKSAPTTPKEPA---PTTIKSAPTTPKEPA-----PTTTK 309
Db 294 PAEAKSPAEPKSPAEVKSAPVAKSPAEPKSPAEVKSAPVAKSPAEPKSPAEVKSAPVAK 352
QY 310 EP-----APTTPKEPA-----PTTTPKEPA-----PTTTPKEPA-----PTTTPKEPA-- 351
Db 353 SPGEAKSPAEPKSPAEVKSAPVAKSPAEPKSPAEVKSAPVAKSPAEPKSPAEVKSAPVAK 411
QY 352 --PTTTPKEPA---PTTTPKEPA---TPTTPKEPAAPTTPKEPA---APTTPKEPA---PTA 393
Db 412 KSPVEAKSPAEPKSPAEVKSAPVAKSPAEPKSPAEVKSAPVAKSPAEPKSPAEVKSAPVAK 470
QY 394 PKKPAPTTPKEPAAPTTPKEPAAPTTPKEPAAPTTPKEPAAPTTPKEPAAPTTPKEPAAPTTP 444
Db 471 AKSPAEPKSPAEVKSAPVAKSPAEPKSPAEVKSAPVAKSPAEPKSPAEVKSAPVAKSPAEPK 528
QY 445 APTTTKS-----APTTPKEPSPTTTPKEPA-----PTTTPKEPAAPTTPKEPAAPTTP 495
Db 529 VKEGAKSLAEAKSPAEPKSPAEVKSAPVAKSPAEPKSPAEVKSAPVAKSPAEPKSPAEVKSAP 588
QY 496 PKEPAPTTPKPAAPTTPKEPAAPTTPKEPAAPTTPKEPAAPTTPKEPAAPTTPKEPAAPTTP 555
Db 589 AKPFAKEEAKRPAIRSPQVKSAPVAKSPAEPKSPAEVKSAPVAKSPAEPKSPAEVKSAPVAK 644
QY 556 APTTPEEPTPTTPKEPAAPTTPKAAAPTTPKEPAAPTTPKEPAAPTTPKEPAAPTTPKEPAAPT 615
Db 645 ---KAKEPPKVEEKTAPKTEVKSAPVAKSPAEPKSPAEVKSAPVAKSPAEPKSPAEVKSAPVAK 700
QY 616 PKGTAPTTLKEPAAPTTPKPAAPTTPKEPAAPTTPKEPAAPTTPKEPAAPTTPKEPAAPTTP 675
Db 701 KKEEA-----KEKKAAPPEEETPAKLVKEAKPKAEADAKA-----KEPSKP 744
QY 676 TPKEPAAPTTPKGTAPTTLKEPAAPTTPKPAAPTTPKEPAAPTTPKEPAAPTTPKEPAAPTTP 735
Db 745 SEKE-----KEEVPAAPEKKTKE-----EKTESKKREKPKMEAK 784
QY 736 TTPKEPAAPTTPKPAAPTTPPETPTTSEVSTPTTTPKEPTTIHKSPDE 782
Db 785 AKED-----KGLPOEPSKPKTEKAESSTTDQKDSQPSKAPED 824

```

Search completed: April 26, 2002, 16:33:31
Job time: 641 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:26:24 ; Search time 61.21 Seconds
(without alignments)
481.977 Million cell updates/sec

Title: AA8
Perfect score: 7054
Sequence: 1 MAWKTLPIYLLLSVFVIQ.....ARATTRSGQTLKVVWYVNC 1311

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	522	7.4	805	US-09-103-429A-4	Sequence 4, Appl
2	508	7.2	1837	US-08-928-361B-5	Sequence 5, Appl
3	506.5	7.2	744	5202236-25	Patent No. 5202236
4	498.5	7.1	786	US-09-103-429A-3	Sequence 3, Appl
5	489	6.9	1721	US-08-700-651-5	Sequence 5, Appl
6	489	6.9	1721	US-08-928-361B-6	Sequence 6, Appl
7	488.5	6.9	826	US-07-638-431-2	Sequence 2, Appl
8	488.5	6.9	826	PCT-US92-00018-2	Patent No. 5202236
9	476.5	6.7	652	5202236-13	Sequence 5, Appl
10	452	6.4	960	US-09-219-849-5	Sequence 5, Appl
11	424.5	6.0	1867	US-08-479-537A-5	Sequence 5, Appl
12	424.5	6.0	1867	US-09-083-116-5	Sequence 5, Appl
13	424.5	6.0	2035	US-08-479-537A-2	Sequence 2, Appl
14	424.5	6.0	2035	US-09-083-116-2	Sequence 2, Appl
15	419.5	5.9	2476	US-08-276-967-2	Sequence 2, Appl
16	419	5.9	1185	US-09-041-886-23	Sequence 23, Appl
17	417	5.9	829	US-08-642-255-132	Sequence 132, Appl
18	417	5.9	829	US-08-397-633A-53	Sequence 53, Appl
19	417	5.9	837	US-08-175-155-68	Sequence 68, Appl
20	417	5.9	837	US-08-477-509B-103	Sequence 103, Appl
21	417	5.9	837	US-08-642-255-101	Sequence 101, Appl
22	417	5.9	837	US-08-707-237A-75	Sequence 75, Appl
23	417	5.9	837	US-08-482-085B-103	Sequence 103, Appl
24	417	5.9	897	US-08-397-633A-50	Sequence 50, Appl
25	413.5	5.9	907	US-08-783-774-2	Sequence 2, Appl
26	413.5	5.9	907	PCT-US95-04611A-19	Sequence 19, Appl
27	404.5	5.7	408	US-07-609-716-65	Sequence 65, Appl

28	404.5	5.7	408	4	US-08-475-411A-65	Sequence 65, Appl
29	404.5	5.7	408	4	US-08-478-029A-65	Sequence 65, Appl
30	398.5	5.6	682	1	US-08-642-255-126	Sequence 126, Appl
31	398.5	5.4	1848	4	US-08-397-633A-36	Sequence 36, Appl
32	381	5.4	1848	4	US-08-296-791-6	Sequence 6, Appl
33	381	5.4	1848	5	PCT-US95-10661A-6	Sequence 6, Appl
34	376	5.3	960	4	US-09-219-849-6	Sequence 6, Appl
35	375.5	5.3	1231	3	US-08-904-263A-4	Sequence 4, Appl
36	374.5	5.3	1537	1	US-08-325-267A-2	Sequence 2, Appl
37	371	5.3	761	2	US-08-707-237A-84	Sequence 84, Appl
38	371	5.3	762	1	US-08-642-255-114	Sequence 114, Appl
39	371	5.3	762	1	US-08-397-633A-26	Sequence 26, Appl
40	370.5	5.2	1064	1	US-08-642-255-62	Sequence 62, Appl
41	368	5.2	762	1	US-08-642-255-120	Sequence 120, Appl
42	368	5.2	762	1	US-08-397-633A-31	Sequence 31, Appl
43	367.5	5.2	1187	1	US-08-320-559-28	Sequence 28, Appl
44	367.5	5.2	1187	3	US-08-545-860D-28	Sequence 28, Appl
45	367.5	5.2	1187	5	PCT-US94-04496-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-09-103-429A-4
; Sequence 4, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang, Ping
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
; TITLE OF INVENTION: CDNA and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
; STREET: 118 No. 6187558th Tioga
; CITY: Ithaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09103,429A
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-2000
; TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 805 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; TISSUE TYPE: peritrophic membrane
; US-09-103-429A-4

Query Match 7.4%; Score 522; DB 4; Length 805;
Best Local Similarity 27.8%; Pred. No. 1.5e-25;

QY 518 -----TTPKETA-PTTP-----KKL 531
Db 842 AGQIADTSLNLPVQTHKSTGLPIDWVGLPDPKSGNLVHYPTNQTMSGLSVSLAAKNL 901
QY 532 T-----PTTPEKLIAPTPE-----545
Db 902 TVDTEYGLPIDTILTYGLDPLDPLVSLPNTPETGLFDPISDEIMNGTIAGIVSGISASES 961
QY 546 ----KPA2PTPEE-----LAPTT-----PEEPTPTP--EEPAPTTPKAAA 580
Db 962 LLSQKSAPIDATTNNVGEFGLLNPAATGVNMPGSLGSEQTPFSPEIDGGIIPPEVAA 1021
QY 581 PNTKREPAPTTPKEPAPTTPKE-----PA 604
Db 1022 ANADKFKLSIPSPESIPEDQKIDSISELMYDIESRLIGOVSKRPIGOSIAGDLNPI 1081
QY 605 PTTKETAPTTPKGTATTLKEPAPTTPKPAKPAKELAPTTTKEPTSTSDKPAAPTTPKGT 664
Db 1082 MKTPTQDSVTKPIDPTT---GLFPNP--PTGHLINPTNNMTDSSFAGAYKVAVNGI 1136
QY 665 APTT----PKEPAPTTPKGTAPTTLKEPAPTTP-----KKPAK-- 707
Db 1137 KTDNVYGLPVDEITGLPKDPVSDIPFNSITGELVDPSTGKPIPNNTAGIVSGKRLGPIE 1196
QY 708 ----ELAPTTTKGP-----TSTT-----SDKPAPTTPKETAPTTPKEPA--- 742
Db 1197 DENGNLFPDSTKLGIDGNLNNLNPENTNSTVSGSTGKPKPGIPVNGGVVDPDEAKDQ 1256
QY 743 -----PTTPKPAAPTTPPTTSEVSTPTTKE--PTTIHKS---DDESTPEL 787
Db 1257 ADKCKDGLVPTTNSINKDPTVNTQYNTTGNINPTGKVIPLGSLPGLNPSFNTPQ 1316
QY 788 SAETPKALE---NSPKPEGVPTTKTPAATKPEMTTAKDKTTERDLRTTPTTAAPKM 844
Db 1317 TDEITGKPDVTVGLPDPSTGEIIDATKLPVGSVAGDEILTEVLNITDEVGLP-I 1375
QY 845 TKETATTEKTESKITATTTQVSTTTQDTPFKITLTKTLTAPKVTTKKTIITTE- 903
Db 1376 DLETGLPRD-----VSLGQLPNGTLVD-----PSNKKRPIGSHS 1411
QY 904 --INNKPEETAAPKDRATNSKATTPKPKP--TKAP--KKPTSTKKPKTMRVRKPKT--- 955
Db 1412 GFINGTSGEOSHEKDPSTG-----KPLDNTGLPDEDSGLNPTGDKLOGSHSGTF 1465
QY 956 TTPPRMTSTMPDELNPTSRIAEAMLTQTTTRPNQTPNSKLVEVNP-K-SEDA----- 1004
Db 1466 MPVPGKPGQGGIMTPEQILEAL-----NKLPTSNVNI SPRSSDAVDPDRPTNWW 1518
QY 1005 -----GGAGETPHMLL-----RPHVFMEVTPMDYLPRVDPNOGII 1042
Db 1519 NKISGQTFQVDGKKTIPGSAASVIHTALGTPTQDPTTGLP--SDPSTGLPFTPGFNVLV 1576
QY 1043 NPMLSD-----ETNIC-----NGKPDVGLTTL--RNGTLV-AF 1072
Db 1577 DPQTGEQIKGSPVSVLYVKEKNIVTEAAYGLPVDPKTGFPIDPISYLPFAKNGELIDPI 1636
QY 1073 RGHVF 1077
Db 1637 SGKIF 1641

RESULT 3

5202236-25

; Patent No. 5202236

; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,

; SUSAN L.; MCCANDLIS, RUSS; WEI, TENA; FILPUILA, DAVID

; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE

; PROTEIN

; NUMBER OF SEQUENCES: 39

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/528,762

; FILING DATE: 25-MAY-1990

; APPLICATION NUMBER: 82,456

; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
; SEQ ID NO:25: 744
; LENGTH: 744
; 5202236-25

Query Match 7.2%; Score 506.5; DB 6; Length 744;
Best Local Similarity 29.9%; Pred. No. 1.3e-24;
Matches 239; Conservative 78; Mismatches 308; Indels 175; Gaps 50;

QY 157 PKITTAKPINPRPSLPNSDTSKETSLSLVNKETTVETKETTNTKOTSTDGKKTTSK 216
Db 24 PKMTYPTTPKPKPSYP-----TYKSKPTY-----KPIIT----- 53
QY 217 TQSIKTSADKLAPTSKVLAKPTPKAETTTKGPAALTTTPKEPTTTPKE-----PASTTPKE 272
Db 54 -----YPTTK--AKPS-----YPTTKPKKTYPTTKPKLTYPPTTKPKP 92
QY 273 PTPTTIKSAPT--TPKEPAPTTPKSAPTTPKEPAPTTPKPAAPTTPKEPAPTTPKEA--P 329
Db 93 SYPTTKSKPTYKPKITYPTTKAKPSYPTTKPKTYPTTKPKTYPTTKPKTYPTTKPKASYP 151
QY 330 TTKSAPTTPKEPAPTTPKPKAPT--TPKEPAPT-----PKEPTTPTPKEPAPTTPKEP--- 381
Db 152 PTYKPKPSYP--PSYKTKTYPTTKPKLTYPPTTKPKPSYPTTKPKPSYPTTKPKTYPTTKPKLT 209
QY 382 -APTTPKEPA--PTAPKPA--PTTPKEPA--PTTPKEPA--PTTPKEPS--PTTPKEPA 430
Db 210 YPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPS 269
QY 431 -PTTKSAPTTPKEPAPTTPKSAPTTPKPSPTTPKEPA--PTTPKEPA-----PTTPKPKP 483
Db 270 YPTTKAKPSYT--YPTTKAKPTY-----PTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPS 322
QY 484 A--PTTPKEPA--PTTPKEPA--PTTKKPA--PTAPKEPAPTTPKETAPTTPKLTPTT 535
Db 323 SYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPS 382
QY 536 PEKLAPTPE--KPAPTTPEELAPTTPPEPT--PTTPKEPA--PTTPKA--AAPNT-----P 584
Db 383 TYKAKPSYPTTKAKPSY-----PTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 438
QY 585 KEPAPTTPKEPAPT-----PKEPAPTTPKETAPT--PKGTAPTTLKEPAPTTPKPKAPK 638
Db 439 SYPTTKPKISYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 498
QY 639 ELAPTTPKEPTSTSDKPAAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPA- 697
Db 499 SYPTTKAKAPTTPSYTKAKPTYK--ARPSY-----PTYKAKPS-----YPTTKAKPSY 546
QY 698 -PTTPKPKAPKELAPTTTKGPT--STTSKPAAPTTPKETAPTTPKEPAPTTPKPA--PT 752
Db 547 PPTYKAKPTYK--AKPTYKAKPTYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTT 601
QY 753 ---TPETPTTSEVSTPTTKE-----PTTIHKSPTD-----ESTPELSAAPT--PKALENS 799
Db 602 YKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAK 661
QY 800 KPEPGVPTTKTAA-----TKPEMTTAKDKTTERDLRTTPTTAAAP-----KMT 845
Db 662 PSYP--PTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 719
QY 846 KETATTEKTESKITATTT 865
Db 720 YKAKPTYPSTYKAKPTYPST 739

RESULT 4

US-09-103-429A-3

Sequence 3, Application US/09103429A
Patent No. 6187558
GENERAL INFORMATION:
APPLICANT: Granados, Robert R
APPLICANT: Wang, Ping
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
TITLE OF INVENTION: cDNA and Related Products and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
STREET: 118 No. 6187558th Tlaga
CITY: Ithaca
STATE: NY
COUNTRY: USA
ZIP: 14850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,429A
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Michaels, Christopher A
REGISTRATION NUMBER: 34,390
REFERENCE/DOCKET NUMBER: BTI-39
TELEPHONE: (607) 256-2000
TELEFAX: (607) 256-3628
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 786 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoplusia ni
TISSUE TYPE: peritrophic membrane
US-09-103-429A-3

Query Match 7.1%; Score 498.5; DB 4; Length 786;
Best Local Similarity 29.1%; Pred. No. 4.6e-24;
Matches 200; Conservative 25; Mismatches 207; Indels 255; Gaps 32;

QY 238 PTPKAETTKGPALTTKPEPTTPKPEASTTKPEPTTTIKSAPTTPKPEAPTTPKTSAP 297
DB 93 PGPFAE-TTQAPAT- - - - -QAPTITQAPTIT- - - - -QAPTITQ- - - - -ATTITQAP 135
QY 298 TTPKEPAPTTPKPEAPTTPKPEAPTTPKTSAPTTPKPEAPTTPKPEAPTTPKPE 357
DB 136 TTITQ-APTITQ-ATTIT- - - - -QAPTITQ-APTITQAPTITQ-APTITQ-APTITQ- 183
QY 358 PAPTTPKEPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPE 414
DB 184 -APTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQ 241
QY 415 - - - - -PTTKEPSPTTPK- 427
DB 242 ELLPNCGRADFDIHLIPHDKYCNLFYQCSNGYTFEQRCEPLFENPVQRCDSANVEC 301
QY 428 - - - - -EPAPTIT- 434
DB 302 DGEISPAFPVTEGNEDEDIDGLDNGCPANFEIDWLLPHGNRCDKYQCVHGNLVERR 361
QY 435 - - - - -KSAPTTPKPEAPTTPKSAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPE 467
DB 362 CGAGTHFSFELQOCDHIELVGLTLPGESESEVDVEDA-CTGWYCPTEPIEWEPLNCP 420

QY 468 APTT- - - - -PKEP- - - - -APTTPKKPAPT- - - - - 487
DB 421 ADFSDHLLPHESDCGYLQCVHGQTIAPPCPNLHFSPATQSCSPVTAGCQVFECDSD 480
QY 488 - - - - -PKEPAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP 544
DB 481 NOCTSTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAA 533
QY 545 EKPAPTPEELAPTTPPEPTTPPEAPTTPPEAPTTPPEAPTTPPEAPTTPPEAPTTPKE 602
DB 534 - - - - -PAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 587
QY 603 PAPTTPKETAPTTPKGTAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPE 662
DB 588 AAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAA 632
QY 663 GTAPTTPKPEAPTTPKPEAPTTPKGTAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 722
DB 633 TAAPT- - - - -AAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 677
QY 723 DKPAPTTPKETAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP 749
DB 678 - - - - -APATTPEDDDIDP- - - - -PLPNDPINP 699

RESULT 5
US-08-700-651-5
Sequence 5, Application US/08700651B
Patent No. 6015882
GENERAL INFORMATION:
APPLICANT: PETERSEN, CAROLYN
APPLICANT: LEECH, JAMES
APPLICANT: NELSON, RICHARD, C.
APPLICANT: GUT, JIRI
TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
FILE REFERENCE: 480.19-4 (HV)
CURRENT APPLICATION NUMBER: US/08/700,651B
CURRENT FILING DATE: 1997-08-14
EARLIER APPLICATION NUMBER: 08/415,751
EARLIER FILING DATE: 1995-04-03
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 1721
TYPE: PRT
ORGANISM: Cryptosporidium parvum
US-08-700-651-5

Query Match 6.9%; Score 489; DB 3; Length 1721;
Best Local Similarity 22.1%; Pred. No. 4.5e-23;
Matches 284; Conservative 82; Mismatches 476; Indels 442; Gaps 41;

QY 111 KKNRTKKKPTKPPVDEAGSLDNGDFKVTTPDTSTQHNKYSTSPKITTAKPINRPS 170
DB 116 RSNETK- - - - -TTEPSANTYAGYRSN-ETKTTEPSANT- - - - -NFLVDPKI- - - - -N 158
QY 171 LPPNSDTSKE- - - - -TSITVNETTETTKETTTTNTKOTSTDGKEKTTSAK 215
DB 159 APCNSENSEGOGLFDMGSKVYIPYTKCVGKHNTTTTTTTTTTTTTTTTTTTTT 208
QY 216 ETQSIEKTSADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKPEAPTTPKPEPT 275
DB 209 - - - - -TT 240
QY 276 TTIKSAPTTPKPEAPTTPKSAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKSA 335
DB 241 TT 300
QY 336 PTTTPKEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTAP 395


```
Db 301 TTTTTPKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT----- 350
QY 396 KPAPTTKPEAPPTPKPEAPPTTKPEAPPTTKPEAPPTTKSAPTITKEAPPTTKSAPTT 455
Db 351 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 406
QY 456 PKE-----PSPPT----- 463
Db 407 TSETSVIKPDWCKLEKNGCEAKGATYGVGVIKDGRIENGMAFTMIPNDTHVRFREK 466
QY 464 -----TKEP-----APTTPKE 474
Db 467 VKDVGNTISVRCRGAGKLEFPDRSLDFTIPPVAGHNSCSIIIVGSGDKIHSVSPYSGKD 526
QY 475 PAPTTPKKPAPTTPKE-----PAPTTPKEAPPTTKKPAAPTAPKEPA 516
Db 527 VSLIS-----APIQSELFNEVYCDCTAKYGAIHSGVQTSADFVTTTAKPTTTT--TGA 580
QY 517 PTPPKETAPPTPKKLPTTPEKLAPTPEKPAPTTPEELAPTTPPEEPTTPPEEAPPTP 576
Db 581 PGQPTTTTGGSPSKPTTTTAKATTT-----TTLNPIITTTTQKPTTT-----TT- 627
QY 577 KAAPNTPEAPPTPKPEAPPTPKETAPTTPKGTAPPTLKEAPPTPKKPA 636
Db 628 --KVPCKPIATTTTLKPIVTTTTKATTTTTTTPPT-----TTTTKRDEMTTTTTPL 680
QY 637 PK-----ELAPTTPKPT-----TTSDKPAPTPKG--TAPTTPKEAPPT 676
Db 681 PDIGDIEITPIEKMLDKYTRMIYDYNGLLSDSDERIPGSOAGIADTSNLFVQTH 740
QY 677 PKPEAPPTPKGTAPTTLKE---PAPTTPKKPAKELAPTTTKGPTSTTSKPAPTPKET 733
Db 741 KSTGLPIDPMVGLPFDPKSGLNVLHPTNQTMSGLSVSYLAAKNLVTDITYG--LPIDT 798
QY 734 APTTPKEAPPTPKKPAPTTPPTTPSVSTPT-----TTKEPTTIHS----- 779
Db 799 LTGYPLDPSLIPFN--PETGELFDISDEIMNGTIAGIVSGISASESLLSKSALIDPA 856
QY 780 -----PDEST-----PELSAETPKALENSPK 801
Db 857 TNMVGEFGLLNPATGVMIPLGFLGSEQTFSEIEDGGIIPPEVAANADKFLSIP- 915
QY 802 EPGVPTTKTPAATKPEMTTAKDKTTER-----DLRTTPTTTAAPKM 844
Db 916 -PSVP-----ESIP-----KQKIDISIELMYDIESGRLIGOVSKRPIPGSIAGDLP 963
QY 845 TKETATTTKTESKITAT--TQVSTTQDTPPKITLTKTTTLAPKVTTKKTTTTE 903
Db 964 IMKPTQTSVTKPIDPTTGLPFNPPTGHLINPTNNMTDSSFAGAYKAYAVSNGIKTN 1023
QY 904 IMNKP--EETAKPKD-----RATNSKATTPKPKPTKAPKPTSTTKPKTMP----- 948
Db 1024 VYGLPVEITGLPKDGSIDPFNSTGELVDPSGTGPIINNSTAGIVSGKPLPIEDENG 1083
QY 949 -----RVKPKTTPTRKMTS--TMPE----- 968
Db 1084 NLPDPSNLPIDGNLNVPNETNSTVSGSTSGTTKPKGIPVNGGVVPDEAKDQADKG 1143
QY 969 -----LNPTSRIAEMLQTTTRNQTNPNSKLVEVNPKSEDAGGAEETHMLLRPHVEMP 1023
Db 1144 KGLIIVPPTINSINKDPVTNTQVSNNTGNI-----INP--ETGKVIPLGSLGSLNYPSTNTP 1197
QY 1024 EVTPD-----MDYLPVPPNPGIINP-----MLSDENINCKG----- 1056
Db 1198 QQTDEITGVPDVTGVLDPDPSGTGEIIDPATKLPIGCVAGDEILFEVLNITDVTGLP 1257
QY 1057 -----PVDGLTTLRNGTLV 1070
Db 1258 IDLETGLPRDPVPSGLPQLPNGTLV 1281
```

RESULT 6

```
US-08-928-361B-6
; Sequence 6, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1721 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-928-361B-6
```

Query Match 6.9%; Score 489; DB 3; Length 1721;
Best Local Similarity 22.1%; Pred. No. 4.5e-23;
Matches 284; Conservative 82; Mismatches 476; Indels 442; Gaps 41;

```
QY 111 KKNRKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDTSTQHNVYSTSPKITTAKPINRPS 170
Db 116 RSNETK---TTEPSANTYAGVRSN-ETKTEPSANT---NFLLVDPKI-----N 156
QY 171 LPPNSDTSKE-----TSLTVNKETTIVETKETTNTKOTSTDGKEKTTSAK 215
Db 159 APCNSENSEQGIQFDMGSKVYIPYTKCVGKHKTITTTTTTTTTTTTTTTT- 208
QY 216 ETQSIKTSKADLAPTSKVLAKPTPKAETTKGPAULTTPKEPTPTTPKEPASTTPKEPTP 275
Db 209 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 240
QY 276 TTIKSAPTTPKPEAPTTTKSAPTTPKEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSA 335
Db 241 TTTTITTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 300
QY 336 PTPPKPEAPPTPKKPAPTTPKEAPTTPKPEPTPTTPKEAPTTTKPEAPTTTKPEAPTAPK 395
Db 301 TTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 350
QY 396 KPAPTTKPEAPPTPKPEAPTTTKPESPPTPKPEAPTTTKSAPTITKEAPPTTKSAPTT 455
Db 351 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 406
```


STREET: P.O. Box 1404
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: United States
 ZIP: 22313-1404
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/083,116
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/479,537
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/FR91/00835
 FILING DATE: 23-OCT-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/039,320
 FILING DATE: 04-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/403,576
 FILING DATE: 14-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Teskin, Robin L.
 REGISTRATION NUMBER: 35,030
 REFERENCE/DOCKET NUMBER: 017753-025
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1867 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 128..1727
 OTHER INFORMATION: /note= "The amino acids spanning 128 to 1727 constitute a repeated region wherein the repeat is 20 amino acids, 17 of which are fixed. The number of such OTHER INFORMATION: repeats varies from 1 to 40."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 134
 OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 144
 OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA or ACG; and Asn = AAT or AAC."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 147
 OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 1..21
 OTHER INFORMATION: /note= "Amino acids 1 to 21 are a 21 amino acid precursor sequence."
 US-09-083-116-5

Query Match 6.0%; Score 424.5; DB 4; Length 1867;
 Best Local Similarity 23.7%; Pred. No. 6e-19;
 Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY	181	TSLTVNKEITVETKETTTTNKQSTDGKEKTTSAKETQSTKTSKAD-LAPTSKVLAKPT	239
DB	16	TVLTV-----VTGSHASSTPGGEKETSATORSSVPSSTERNNAVMTSSVLSHS	65
QY	240	P-KAETTTTKGP--ALTTPKP-----TPTTPKEPASTTPKEPTTTTIKSP	282
DB	66	PGSGSSTTQGDVTLAPATEPASGAATWGDVTSVPTRPALGSTTTPPAHDVT--SAP	122
QY	283	TTPKPAPPTT-----KSAPTTPKP-----APTTPKEPATTTPKEPAPT	323
DB	123	--DNKPAPGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT	180
QY	324	TKP--PAPTTT-----KSAPTTPKPAPTTPK-----KPAPTTPKPAPTTPKEPTPTT	370
DB	181	APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT	240
QY	371	PKE--PAPTTKPE-----APTTPKEPAPTAPK-----KPAPTTPKPAPTTPKEPAPT	417
DB	241	APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT	300
QY	418	TKPSPTTPKEPAPT--TKSAPTTTKEPAPT-----TKSAPTTKPEPSPTTK-----E	466
DB	301	APDXRP--XPGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT	359
QY	467	PAPTTPKEPAPTTPK-----KPAPTTPKPAPTTPK-----EPAPTTTTPKAPTAPKEPAPT	518
DB	360	SAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT	419
QY	519	TPKETAP-----TTPPKLTPTTPEKLAAPTTPKEPAPTTPPELAPTTPPEPTPTTPEEPAPT	574
DB	420	SAPDXRXPXGSTAPXAHGVT-----APDXRXPXGSTAPXAHGVTAPDXRXP-----PGST	471
QY	575	TPKA---AAPNTPKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKGT--APTTLKEPAP	629
DB	472	APXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXP--XPGSTAPXAHGVTAPDXRXPXGS	530
QY	630	TPP-----KKPAPKELAPTTTKEPTSTTSDKAP--TTPKGTAPTTPKEPAPTTP	677
DB	531	TAPXAHGVTAPDXRXPXGSTAP--XAHGVTAPDXRXPXGSTAPXAHGVT-----APDXR	585
QY	678	KEPAPTTPKG-----TAPTTLKEPAPTTP-----KKPAPKELAPTTTKEPTSTTSD	723
DB	586	PXPGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAP--XAHGVTAPDX	644
QY	724	KPAP--TTPK-----ETAPTTPKEPAPTTP-----KKPAPTTPETPTTSEVSTP	767
DB	645	RPXPGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGS--TAPXAHGVTAP	702
QY	768	TTTKEPT-----IHKSPESTPELSAPTPKALENSPKPVGPTTKTPAA-----	813
DB	703	DXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAP	762
QY	814	-TKPEMTTAKDK---TTERDLRTTPTTAAPKMTKATATTTTEKTTESKITATTQVTS	869
DB	763	DXRXPXGSTAPXAHGVTAPDXRXPXGST--APXAHGVTAPDXRXPXGSTAPXAHGVT	820
QY	870	TTTQDTPPKITTLKTTTLAPKVTTTKKTTTTTTEIMKNPEETAKPKDRATNSKATTPKPQ	929
DB	821	APDXRXPXG-----STAPXAHGVT-----SAPDXRXPXGSTAPXAHGVTAPDXRXP	868
QY	930	KPTKAPKKPTSTKKPTMPVRVKPTTPTPKMTSTMPELNPTSRIAEAMLQTTTRPNOT	989
DB	869	-GSTAPXAHGVTAPDXRXP--XPGSTAPXAHGVTSA--PDXRXP-----PGST	911
QY	990	-PNSKLVEVNPKSEDAEGEETPHMLLRPHVFMPEVTPDMXYLPRVPMNOGIINPMUSD	1048
DB	912	APXAHGVTAPDXRXPXGSTAPXAH-----GVTAPDXRXPXGSTA---PXAHG	957
QY	1049	ETNICNCKPVDGLTT-LRNGTLVAFRGHYFWMLSFSPFPSPARRITEWVGIPSPID	1103

Db 958 VTSAPDXRXPXGSTAPXAHGVTSAPD-----PDXRXPXGSTAPXAHGVTSAPD 1003

RESULT 13

US-08-479-537A-2

Sequence 2, Application US/08479537A

Patent No. 5861381

GENERAL INFORMATION:

APPLICANT: CHAMBER, Pierre

APPLICANT: KIENY, Marie-Paule

APPLICANT: LATHE, Richard

APPLICANT: HAREUVENI, Mara

TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE

TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/479,537A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 90/13101

FILING DATE: 23-OCT-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/FR91/00835

FILING DATE: 23-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/039,320

FILING DATE: 04-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/403,576

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 017753-025

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2035 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Peptide

LOCATION: 128..1899

OTHER INFORMATION: /note= "The amino acids spanning

OTHER INFORMATION: 128 to 1899 constitute a repeated

OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such

OTHER INFORMATION: repeats varies from 1 to 40."

FEATURE:

NAME/KEY: Peptide

LOCATION: 134

OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa

OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,

OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 144

OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa

OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,

OTHER INFORMATION: or ACG; and Asn = AAT or AAC."

FEATURE:

NAME/KEY: Peptide

LOCATION: 147

OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa

OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,

OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 1..21

OTHER INFORMATION: /note= "Amino acids 1 to 21 are a

OTHER INFORMATION: 21 amino acid precursor sequence."

US-08-479-537A-2

Query Match 6.0%; Score 424.5; DB 2; Length 2035;

Best Local Similarity 23.7%; Pred. No. 6.6e-19;

Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 181 TSLTVNKETTVEKTTTNNKQSTDKGKETTSAKETQSIKTSKAD-LAPTSKVLAKPT 239

Db 16 TVLTV-----VTGSGHASSTPGCEKETSATQSSVPSSSTERNASMTSSVLSHS 65

QY 240 P-KAETTTKGP--ALATPKP-----TPTTPKEPASTTPKEPTPTTIKSAP 282

Db 66 PGSGSSTTQGDVTLAPATEPASGAATWGQDVTSPVTRPALGSIITPAHDVT---SAP 122

QY 283 TTPKEPAPTTT-----KSAPTTKPEP-----APTTTKEPAPTTTKEPAPTT 323

Db 123 --DNKPAGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 180

QY 324 TKE--PAPTTT-----KSAPTTKPEPAPTTPK---KPAPTTKPEPAPTTPKEPTPTT 370

Db 181 APDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 240

QY 371 PKE--PAPTTKPEP-----APTTTKEPAPTAPK---KPAPTTKPEPAPTTPKEPAPT 417

Db 241 APDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 300

QY 418 TKEPSTTPKEPAPT--TKSAPTTTKEPAPT-----TKSAPTTKEPSPPTTK----E 466

Db 301 APDXR--XPGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 359

QY 467 PAPTTKPEPAPTTPK---KPAPTTKPEPAPTTPK---EPAPTTTTPKAPAPKEPAPT 518

Db 360 SAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 419

QY 519 TPKEAP-----TTPKKLTPTTPEKLAPTTPEKAPAPTTPELAPTTPEEPTTPPEPAPT 574

Db 420 SAPDXRXPXGSTAPXAHGVTS-----APDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTS 471

QY 575 TPKA---AAPNTPKPEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKGT--APTTLKEPAP 629

Db 472 APXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXP--XPGSTAPXAHGVTSAPDXRXPXG 530

QY 630 TTP-----KKPAPKELAPTTTKEPTSTTSKAP--TTPKGTAPTTTKEPAPTTP 677

Db 531 TAPXAHGVTSAPDXRXPXGSTAP--XAHGVTSAPDXRXPXGSTAPXAHGVTS-----APDXR 585

QY 678 KEAPAPTPKG---TAPTTLKEPAPTTP-----KKPAPKELAPTTTKEPTSTTS 723

Db 586 XPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAP--XAHGVTSAPDX 644

QY 724 KPAP--TTPK---ETAPTTKPEPAPTTP-----KKPAPTTPEPTTSEVSTP 767

Db 645 RXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXG--TAPXAHGVTSAP 702

QY 768 TTTTKEPT-----IHKSPDESTPELSAETPTPKALENSPKPEGVPTTKTPAA----- 813

Db 703 DXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAPDXRXPXGSTAPXAHGVTSAP 762

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:18:41 ; Search time 82.85 Seconds
(without alignments)
890.006 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1167

Perfect score: 5155

Sequence: 1 VKDNKNRTRKKKPTPKPPV.....GKPDVGLTLRLNGTLVAFRG 968

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

PIR_68: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1062.5	20.6	3020	A43932	mucin 2 precursor,
2	942	18.3	1664	T18262	S-layer protein -
3	853	16.5	1489	T31108	cyst germination s
4	802	15.6	1274	T16251	hypothetical prote
5	784	15.2	1188	S49915	extensin-like prot
6	776	15.1	2187	T30826	nascent polypeptid
7	763	14.8	1367	S48478	glucan 1,4-alpha-g
8	671	13.0	1151	T18535	high molecular mas
9	667	12.9	1344	A35175	mucin 1 precursor,
10	659.5	12.8	1229	T25697	hypothetical prote
11	633	12.3	489	T11622	extensin class 1 p
12	632.5	12.3	7962	T13846	elastic titin - hu
13	632	12.3	761	C84672	hypothetical prote
14	629	12.2	3507	T34513	hypothetical prote
15	626.5	12.2	990	T15168	nucleolar phosphop
16	625	12.1	6642	T29757	protein UNC-89 - C
17	624.5	12.1	839	T75518	hypothetical prote
18	620.5	12.0	971	T19431	hypothetical prote
19	617	12.0	3570	T245025	mucin MUC5B, trach
20	607.5	11.8	801	T29018	hypothetical prote
21	607.5	11.8	924	T27923	gene LF3 protein -
22	605	11.7	379	S50125	larval glue protei
23	559.5	10.9	350	T22456	hydroxyproline-ric
24	556.5	10.8	856	T16543	hypothetical prote
25	556.5	10.8	875	S23760	polyphenolic adhes
26	555.5	10.8	1630	T25377	ascites sialoglyco
27	551	10.7	620	T06733	hydroxyproline-ric
28	549	10.6	2232	T34434	hypothetical prote
29	545	10.6	873	A47283	calphotin - fruit

30	542	10.5	369	2	S20500	hydroxyproline-ric
31	538.5	10.4	416	2	JU0465	extensin precursor
32	533	10.3	1087	1	QFMSH	neurofilament trip
33	532	10.3	756	2	T27642	hypothetical prote
34	530.5	10.3	1162	2	JH0557	exo-alpha-sialidas
35	530	10.3	865	2	A47282	calcium-binding pr
36	522	10.1	328	2	JQ0985	hydroxyproline-ric
37	518	10.0	813	2	S70795	vsaA protein precu
38	518	10.0	866	2	T45462	membrane glycoprot
39	518	10.0	1072	1	A37221	neurofilament trip
40	517.5	10.0	1459	2	T32271	hypothetical prote
41	512.5	9.9	867	2	T45463	membrane glycoprot
42	509.5	9.9	662	2	A45155	mucin FIM-C.1 - Af
43	505	9.8	700	2	A54641	interspersed repea
44	504.5	9.8	606	2	A43427	neurofilament trip
45	504	9.8	1832	2	T31113	mucin-like glycopr

ALIGNMENTS

RESULT 1

A43932

mucin 2 precursor, intestinal - human (fragments)

N:Alternate names: mucin SMUC-41

C:Species: Homo sapiens (man)

C>Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999

C:Accession: A49963; A45106; B45106; B33532; A61257; PQ0328; PQ0329

R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.

J. Biol. Chem. 269, 2440-2446, 1994

A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t

A:Reference number: A49963; MUID:94132002

A:Accession: A49963

A:Molecule type: mRNA

A:Residues: 1-639 <GU1>

A:Cross-references: GB:L21998

R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.

J. Biol. Chem. 267, 21375-21383, 1992

A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up

A:Reference number: A45106; MUID:93016075

A:Accession: A45106

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 626-1895 <GU2>

A:Cross-references: GB:M94131; MID:gl86395; PIDN:AAA59163.1; PID:gl86396

A>Note: sequence extracted from NCBI backbone (NCBIP:116706)

A:Accession: B45106

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Cross-references: GB:M94132; MID:gl86397; PIDN:AAA59164.1; PID:gl86398

A:Experimental source: colon

A>Note: sequence extracted from NCBI backbone (NCBIP:116698)

R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,

J. Clin. Invest. 88, 1005-1013, 1991

A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym

A:Reference number: A43932; MUID:91358717

A:Accession: A43932

A:Molecule type: DNA

A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>

A:Cross-references: GB:M74027; MID:gl88863; PIDN:AAA59875.1; PID:gl88864

A>Note: sequence inconsistent with the nucleotide translation

A>Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)

R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.

J. Biol. Chem. 264, 6480-6487, 1989

A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evl

A:Reference number: A33532; MUID:89197956

A:Accession: B33532

A:Molecule type: mRNA

A:Residues: 1916-2193 <GU4>

A:Cross-references: GB:M22405; MID:gl88873; PIDN:AAA36334.1; PID:gl88874

A:Experimental source: intestine

R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

J. Clin. Invest. 87, 77-82, 1991
A:Title: Human bronchus and intestine express the same mucin gene.
A:Reference number: A61257; MUID:91086481
A:Accession: A61257
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
A:Experimental source: bronchus
R.Xu, G.; Huan, L.; Khatir, I.; Sajjan, U.S.; McCOOL, D.; Wang, D.; Jones, C.; Forstner, Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminus
A:Reference number: PQ0328; MUID:92198477
A:Accession: PQ0328
A:Molecule type: mRNA
A:Residues: 2328-2468 <XUG>
A:CROSS-references: GB:M86523
A:Experimental source: small intestine
A:Accession: PQ0329
A:Molecule type: protein
A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
C:Genetics:
A:Gene: GDB:MUC2
A:CROSS-references: GDB:120203; OMIM:158370
A:Map position: 11p15.5-11p15.5
C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von Willebrand factor type C repeat homology
C:Keywords: glycoprotein; intestine; tandem repeat
F:2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 20.6%; Score 1062.5; DB 2; Length 3020;
Best Local Similarity 29.2%; Pred. No. 4.2e-38;
Matches 342; Conservative 74; Mismatches 442; Indels 315; Gaps 39;

QY 32 FKVTTPDTSTTOH---NKVSTPKITAKPINPRSLPNSDTSKETSITVNKETTVEK 88
Db 1240 WEICGPNVGEKHFNCISITRPSLTITTTITLPTP-----TSFTTTTITPTPS 1291
QY 89 ET---TTTNK-----QTSDGKEKTS--AKETQSIEKTSKAD-----121
Db 1292 STVLSTTPKLCCLWDINEDHPSGSDGDRFPGVCGAPEDIECRSVKDPHLSLEQH 1351
QY 122 -----LAPTSKVLAKPTPKAE 137
Db 1352 GQKQCDVSVGFICKNEDQFGNGFLCYDKIRVNCWPMDCIITPSPPTTTPSPPTT 1411
QY 138 TTTKGPALTTPKPTTTPKEPASTTPKEPTTIKSAP--TTPKEPAPTTPKSAPTTPKE 196
Db 1412 TTTTLPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPS 1471
QY 197 PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAP--TTPKEPAPTTPKKPAPTTPKEPAPT 255
Db 1472 P-PTTTPSPPTTTPSPPT--TTTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPSPTT 1529
QY 256 TPKEPTTPKEPAPTTPKEPAPTTPKEPAPTAPKKAPITPKPAPTTPKEPAPTTPKEP 315
Db 1530 TTTTPPTTTPSPPTTTPITP-PTSTTTTLPTTTPSPPTTTPPTTTPPTTTPSPPTTTPS 1588
QY 316 SPPTTPKEPAPTTPKSAP--TTTKPAPTTPKSAP--TTPKEPAPTTPKEPAPTTPKEPAPT 373
Db 1589 PTTTTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSPTT 1648
QY 374 PKKPAPTTPKEPAPTTPKEPAPTTPKKPAPTAPKEPAPTTPKETAPTTPKKLPTTPEKL 433
Db 1649 TTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPSSPTTTP--SPPTT-----TMTTPS-- 1700
QY 434 APTTPKEPAPTTPKEPAPTTPPTTTPKEPAPTTPKEPAPTTPKAAAPTTPKEPAPTTPKEPAPT 493
Db 1701 PTTTPSPPTTTPPTTTPSSTTTP--SPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSSPLTTT 1759
QY 494 PKPAPTTP-----KETAPTTP-----KG-----512
Db 1760 PLPPTTTPPTTTPPTTTPPTTTPCPLCNWGLDGSKPNFKPGGDELIGDVGCGWAA 1819

QY 513 -----512
Db 1820 NISCRATMYDPVPIGOLGOTVVCDVSVGLICKNEDQKPGGVIPMAFCLNVEINVQCECV 1879
QY 513 TAPTTLKEPAPTTPKKPAPKELAPT--TKEPTSTSDKP-----APT--TPKGTAPTTPKE 565
Db 1880 TQPTTM---TTTTTENPTPTTITTTTTPPTTPTSTOSPNGLQAPTPTTITTTTTPPT 1936
QY 566 PAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTPKGTST--TSDKPAPTTP 624
Db 1937 PPTGTQPT--TPTTITTTTTPPTTGTQPTTTLVLTITTTTTPPTTPTSTKSTTTP 1995
QY 625 KETAPTTPKEPAPTTPKKP-----APTPTPTPTTTPPTTTPPTTTPK-----EPTTIHK 672
Db 1996 ITTTTTPVATPTPTGTQPTMIPISITTTTTPPTTPTTGTSTGPTTHTSTAPIAELTISNP 2055
QY 673 SPDESTPELSAEPKALENSPKPCVPTTKPAATKPEMTTAKDKTTERDLRTPEPT 732
Db 2056 PPESSTPQTSRSTSSPLTETSTLLSTLPAIEMTSTAPSTTPTATTTSGGHTLSPPT 2115
QY 733 TAAPKMTKETATTTETKTESKITATTTQVTTSTTTQDTPPTPKITT---LKTTLAPKVTTP 789
Db 2116 TTSPPCTPTRGTT--GSSSAPTPTSVQTTTSATWPTTPTLSTPSIIRTGLRPVSSV 2173
QY 790 KKTITTEIMNKEE-----TAKPKDRATNSKATTP 820
Db 2174 LICCVLNDYYAPGERVYNGTYGTCYFVNCSLSCTLEFYNWSCPSTPTPTPSK--STP 2232
QY 821 KPOKPTKAPKPTSTKPKTTPRVKPKTTPTRKMTSMPELNPTSRILAEAMLQTTTPR 880
Db 2233 TTSKPSSTSKPTPTGTPKPECDFDPPR-----QENETWILCDCEM--ATCKY 2278
QY 881 NOTPNSKLVENP-----KSDAGG-----AEGTTPHMLL-----910
Db 2279 NNTVEIVKVECEPPPPMPTCSNGLQPVVRVEDPDGCGWHWCDCYCTGWD--PHVYTFDGLY 2337
QY 911 -----RPHVFPEVTPDMDYLPVRVNOGIIN 937
Db 2338 YSYQGNCTYVLVEISPSVD-----NFGYIID 2364

RESULT 2
T18262
S-layer protein - Clostridium thermocellum
C:Species: Clostridium thermocellum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18262
R:Fujino, T.; Bequin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose.
A:Reference number: Z18847; MUID:93209931
A:Accession: T18262
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1864 <FU>
A:CROSS-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841.1

Query Match 18.3%; Score 942; DB 2; Length 1664;
Best Local Similarity 33.0%; Pred. No. 3.2e-33;
Matches 276; Conservative 83; Mismatches 312; Indels 166; Gaps 43;

QY 128 VLAKPTP-KAETTKGPALTTPKEPTTPKEPASTTPKEPTTTPKAPTTPKAPTTP 186
Db 758 VVLOPAPKASDEPIPTDTPSDEPTPS-----DEPTPSDEPTPSD 804
QY 187 TKSAPTTPKEPAPTTPKEPAPTTPKEPA-----PTTTPKEPAPTTPKAPTTPKEAPTTPK 242
Db 805 EPTPSTPEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSETPEPIPTDTP 864
QY 243 KPAPTTPKEPAPTTPKEPTTPKAPTTPKEPAPT--TPKEPAPTAPKKAPTTPKEPAPT 301

A:Accession: A26877
A:Molecule type: DNA
A:Residues: 1-242 <YAM>
A:Cross-references: EMBL:M16164; NID:gl72522; PIDN:AAA35014.1; PID:gl72525
A:Accession: B26877
A:Molecule type: DNA
A:Residues: 762-1331 <YA2>
A:Cross-references: EMBL:M16165; NID:gl72523; PIDN:AAA35015.1; PID:gl72526
R:Pardo J.M.; Ianez E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Saccharomyces cerevisiae
A:Reference number: S27281; MUID:89031230
A:Accession: S27281
A:Molecule type: DNA
A:Residues: 1-31 <PAP>
A:Cross-references: EMBL:X13857; NID:q4551; PIDN:CAA32069.1; PID:q4552
R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.G.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A:Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohyphal growth of *Candida albicans*
A:Reference number: JG6123; MUID:96323237
A:Accession: JG6123
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1367 <LAM>
A:Cross-references: GB:U30626; NID:gl1304386; PIDN:AAC49609.1; PID:gl1304387
C:Genetics:
A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A:Cross-references: MIPS:Y1R019c; SGD:S0001458
A:Map position: 9R
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F:5-21/Domain: transmembrane #status predicted <TM1>
F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 14.8%; Score 763; DB 1; Length 1367;
Best Local Similarity 28.5%; Pred. NO. 1.1e-25;
Matches 286; Conservative 98; Mismatches 465; Indels 154; Gaps 43;

QY 35 TPPTDSTTHQNKVSPKTTAKPKNP-----RPLPPNDSRDTSKLTVNKEITVE 86
DB 236 STSESSTTSSTSSSTSSSTAPATPTTCTCKEPPPTTSCTEKPTPHDHTTPC 295
QY 87 TKETTTTNQTSDGKEKTSKAKTSQIEKTSKADLAPTSKVLAKPTPKAET--TKGPA 144
DB 296 TKKKTTSK-TCT---KKTTVPVTPS-SSTESSAPV-----PTPSSSTESSAPV 344
QY 145 LTPKE-----PTTTPKEPASTTPKEPTTIKSA-----TTPKEPATTTKSAPTTP 194
DB 345 TSTSTESSAPVPTPSSSTESSAPVSTSTESSAPVSTSTESSAPVTPSSSTTE 404
QY 195 KEPAPTTTKEPAPTTKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPA 254
DB 405 SSAPVTS---TTESSAPVTS---STTESSAPV---SSTESSAPVSTSTES 453
QY 255 TTPKEPTP---TTPKEPATTTKEPAPTTKPAPTTKKPAPTTKKPAPTTKKPA 307
DB 454 SSAPVTPSSSTESSAPV---SSTESSAPV-VPTPSSSTESSAPVSTSTESS 509
QY 308 APPTTKPSPTPKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTK 363
DB 510 APVTPPS-SSTESSAPAPTSSSTESSAPVSTSTESSAPVSTSTESSAPVPTP 568
QY 364 ----TTPKEPAPT-TPKKPAPTTPKEPAPTTPKEPAPTTPKPAPTTKKPAPT 418
DB 569 VTSTSTESSAPVPTPSSSTESSAPVPT---PSSSTESSAPAPTTPSSSTESS 624
QY 419 PTPPKLPTTPKEKLAPTTPKEKLAPTTPKEKLAPTTPKEKLAPTTPKEKLA 472
DB 625 PVT---SSTESSAP-VPTPSSSTESSAPVPTPSSSTESSAPVPTPSSSTESS 679
QY 473 AAPNTPKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 531

DB 680 SAPVT---SSTESSAPVT---SSTESSAPVPTPSSSTESSAPVPTPSSSTTE 731
QY 532 KELAPTTTKEPTSTSDKPAPTTKGTAAPTTPKEPAP---TTPKEPAPT-TPKGTA 587
DB 732 SSSAPVTPPS-SSTESSAPVSTSTESSAPVPTPSSSTESSAPVPTPSSSTTE 790
QY 588 KEPAPTTPKKPAKELAPTTTKGPTSTTSDKPAPTTKKPAPTTKKPAPTTKKPA 642
DB 791 SAPVPTPSSSTESSAPVPTPSSSNTSSAPSTPSSSTESSAPVPTPSSSTTE 850
QY 643 PAPTTPETPTPTTSVSTPT-----TTPKEPTIHKSPDESTPELSAETPKALE 697
DB 851 SAPVSSSTESSAPVPTPSSSNTSSAPSSI---PFSSTESTSTGT-TVTPSSSK 906
QY 698 GVTPTTKTAAKPEMTTAAKDKTTERDLKTPETTTAAAPKMTKETATTEKTES 757
DB 907 GSQETSVSSSTETIVTKTITTSVTPTTITTVTSTGTSNAGETSGSPKTVIT 966
QY 758 --TTQVSTTTQDTPFKITTLTKTLAPKVT---TKKTIPTT-EIMNKPEETA 811
DB 967 VPTTTTTSVTSSTTITTTVCSTGTSNAGETSGSPKTIITTVPCSTSPSETA- 1021
QY 812 ATNSKATTPKPKP-----TKAPKKPTSTKKKPMRVRKPTTPR 854
DB 1022 ---SESTTTPPTVTVVSTTVVTEYSTKPGGEITTEVTXNIPPTTLTIAPT 1077
QY 855 KMTSTMPELNPTSRIAEMLOTTTRPNQTPNSKLVEVNPKSEDAGAGETPHMLR 914
DB 1078 -SVTVTVNFTPTT-TTIVVCSGT-----NSAGETSGCSPKTVTVPCST 1122
QY 915 FMPEVTPDMYLPVNPNGIINMLSDETNTCNKQPVDTLT 957
DB 1123 GRGEYTTAATLVTAATVTTVVTTESSTGTSNAGETSGTNSA-GKTTTGYTT 1164

RESULT 8
TI8535
high molecular mass nuclear antigen - chicken (fragment)
C:Species: Gallus gallus (chicken)
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: TI8535
R:Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of ch
A:Reference number: Z18955; MUID:9803440
A:Accession: TI8535
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1151 <SHI>
A:Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BAA24137.1

Query Match 13.0%; Score 671; DB 2; Length 1151;
Best Local Similarity 25.9%; Pred. No. 7.3e-22;
Matches 284; Conservative 100; Mismatches 419; Indels 292; Gaps 55;

QY 14 TPKP-----EVDVAGSGDLNGDFKVTPTDSTTHQNKVTS--PKITTAKPINRPS 64
DB 113 TPPPSSQGPAGTTPPSOGAAGAPKGDGTAQPSGTSKSGADGKPAQDVPKATTA-ATEARD- 170
QY 65 LPNDSSTSKETSLTVNKEITVETKTTTNNK--QVSTDGKEKTSKAKTSQIEKTSK 122
DB 171 -----ASAASPTVPKATAEATAVTAASQSAPKAAATAAAVTA--SOSAPKATV-EV 219
QY 123 APTSKVLAKPTEKAEITTKGPAITTPKEPTP-TTPKEPASTTPKEP-----TPTTIKSAPT 177
DB 220 KPAAAVAKEAKAVTAAAPKATAEAKPAPVSTPIPCSSAEAKPLTAASATASKA--T 277
QY 178 TPKEPAPT-----TKSAPTTPKEPAPTTPK----- 203
DB 278 AEAKVPVATASLMATKVTAEAKPAPSPVPKATTDTKAVTATATAGDPVKAAYCABA 337
QY 204 EPAPTTKPEPATTTPKEPATTTKSAPTI-----PKEPA-----PTTPKKPAPTTPK 250

338 KPAPPPTPOOLKAAAAAATGTGLKAPATAPPHGSPRANSHTVTVTPNNVPRAAAATVP- 396
251 EPAPPTPKPTPTT---PKEPAPTTKEPAPTTKPEPAPTPAPKPKAPPTPKPEPAPTTPE 306
397 -TAGAVPKASTGTTAAAPOQVP---KAPVTPPSQQAQVPRAAATAA---APVTPOQ 448
307 P---APTTTKEPSPTTKPEPAPTTKSAPTTTKEPAP---TTTTKSAPTTKPEPSPTTKE 360
449 PVTKAATTTNATPPQIPKAAATTTTATPVTPOQIPKAGTDAAPPAVPKASDGRAAT 508
361 P---APTTTKEPAPTTKPKAPTTKEPAP---TTTTKEPAPT---TTTKPEPAPT---TTTKPEPAPT 406
509 PGVNAATDPKPPPTQSPVSAVTEPKPKQPRAAAPPSNEATPAVPSPSNLKSLPTTIP 568
407 K-----EPAPTPK---ETAPTTKLTPTTPEKLAPTTPEKPAPTTPEELA----- 450
569 KPVLMLALTPQPVTAQWVQLAAKPSPIVPKASPK-ALMTPPPPPPGPRALAAAALLG 627
451 -PTTP-----EPTPTTP---EPAPTPP-----KAAAPNTPKE----- 480
628 LPSSPVASAMHAKVTPRLPASVPVMAASPASGLGPDAAARVALATNAASPGARPEAAGNG 687
481 ----PAPTTKEPAPTTKPEPAPTTKPEAPT-----TPKGTAPT----- 516
688 TLMAPGAANTOMAPICGAAGAAQTAPMGAAHTHVSWMGAGGATQMSPTGAANTHMSPIGA 747
517 -----TLKEPAPTTKPKAPKELAPTTTKEP-----TSTTSKPKAPPTP 555
748 GGAQTQSPMGAAANTQSPMGATTTQSPMGAAATQSPMGAAATQVATSAAGTMOQVSP 807
556 KG---TAPTTTKEPAPTTKPEPAPTTKGTAPTTKBP---APTTPKPKAPKELAPTTTKEP 611
808 MGAATPPQTPSVGAATTP-QPSPM---GAATTLMSPMGAATTPQ---PSPMGAVTTQPP 859
612 ----TSTTSKPK-APTTPKET---APTTPKEP---APTTPKPK-APTTPETPP----- 652
860 PMAATNTTQPPMAASTPOSTPMGAATTTQSPMGATTTQSPMGASTPQAPPTVAGSPT 919
653 PTTSEVSTPTTKEPTTIHKSDESTPEL--SAEPTPKALENSKPEPGVPTTKTAPATKP 710
920 PPPPIPPSTAGTSOPMSKSPPPDPKAPSAQAATSPAAHVANASPGV-TAVSPA---p 975
711 EMTTAKKTDRLRTTETTTAAPKMT-KETATTTEKTTESKIPATTTQVTTSTTQDT 769
976 IGVTASPSADGARLSPGTAATDGGKASPAATADVTEAATD--VTAATAVPA-----EA 1029
770 TPEKITLTTTLAPKVTTKKTTTITTEIMNKPETEAKPKDRATNSKATTPKPKPTKAP 829
1030 AP-----TKAKRSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSPASPA--- 1077
830 KKPTSTKKPKTMRVRKPKTTTTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKL 889
1078 -----PVGDDGQQQMTPGNAQSVPP-----VTEAAVQ----- 1104
890 EVNPKSEDAGGAEGE 904
1105 EAAAAAAGAAE 1119

RESULT 9

A:Accession: A35175

A:Molecule type: mRNA

A:Residues: 1-952,1033-1344 <LIG1>
A:Cross-references: GB:J05288; GB:J05288; NID:9182121; PIDN:AAA35804.1; PID:9182124;A:Experimental source: splice form A
A:Note: Genbank entries HUMEPISIA1 and HUMEPISIA2 present only the amino-and carboxyl

A:Accession: B35175

A:Molecule type: mRNA

A:Residues: 1-19,29-952,1033-1344 <LIG2>
A:Cross-references: GB:J05288; GB:J05288; NID:9182126; PIDN:AAA35806.1; PID:9182129;

A:Experimental source: splice form B

A:Note: Genbank entries HUMEPISIB1 and HUMEPISIB2 present only the amino-and carboxyl

R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhlig, T.; Peat, N.; Burt

J. Biol. Chem. 265, 15296-15299, 1990

A:Title: Molecular cloning and expression of human tumor-associated polymorphic epith

A:Reference number: A35886; MUID:90368715

A:Accession: A35886

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-19,29-992,1033-1344 <GEN>
A:Cross-references: GB:J05581; NID:9188869; PIDN:AAA59876.1; PID:9188870

A:Note: Genbank entry HUMUCAB includes one copy of the tandemly repeated sequence

R:lan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.

J. Biol. Chem. 265, 15294-15299, 1990

A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.

A:Reference number: A35887; MUID:90368716

A:Accession: A35887

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
A:Cross-references: GB:J05582; NID:9189598; PIDN:AAA60019.1; PID:9189599

A:Note: Genbank entry HUMPANMU contains four fewer copies of the tandemly repeated se

R:Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky,

Eur. J. Biochem. 189, 463-473, 1990

A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may gen

A:Reference number: S10571; MUID:90276413

A:Accession: S10572

A:Molecule type: mRNA

A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>
A:Cross-references: EMBL:X52229; NID:937053; PIDN:CAA36478.1; PID:937054

A:Reference number: S40293

A:Accession: S40293

A:Molecule type: mRNA

A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <A>
A:Cross-references: EMBL:X52229; NID:937053; PIDN:CAA36478.1; PID:937054

A:Reference number: S40293

A:Accession: S40293

A:Molecule type: mRNA

A:Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>
A:Cross-references: EMBL:M31823; NID:9181542; PIDN:AAA35757.1; PID:9181543

R:Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu

J. Biochem. 112, 609-615, 1992

A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut ag

A:Reference number: JX0235; MUID:93123189

A:Accession: PX0066

A:Molecule type: mRNA

A:Residues: 998-1011,'ES',1014-1017,1018-1032,'T',1034-1037;1038-1057 <MAS>
A:Experimental source: gastric carcinoma cell

R:Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.

FEBS Lett. 356, 130-136, 1994

A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytoKin

A:Reference number: S51026; MUID:95080414

A:Contents: annotation

A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region

partial repeats. The repeat shown is defined by SmaI nuclease sites.

C:Comment: Serine and threonine residues in the tandem repeat domain are extensively

C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48

C:Genetics:
A:Gene: GDB:MUC1; PUB
A:Cross-references: GDB:120705; OMIM:158340
A:Map position: 1q21-1q23
A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
C:Superfamily: polymorphic epithelial mucin
C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphism
F:1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>
F:1-62/Region: mucin 1 amino-terminal non-repetitive
F:1-23/Domain: signal sequence #link PREA #status predicted <SIGA>
F:1-19-29-32/Domain: signal sequence #link PREB #status predicted <SIGB>
F:1-19-29-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>
F:1-19-29-212-1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice form
F:138-1017/Region: 20-residue repeats (GSTAPPAHGVTSAPDTRPAP)
F:1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
F:1245-1272/Domain: transmembrane #status predicted <TRM>
F:1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 12.8%; Score 667; DB 1; Length 1344;
Best Local Similarity 27.8%; Pred. No. 1.2e-21;
Matches 305; Conservative 77; Mismatches 459; Indels 256; Gaps 56;

QY 77 LTVNKETTVEKTTTNNK--QTSTGCKKTTSAKETQSIKTSKAD-LAPTSKVLAKPT 133
Db 15 LTVLTATAPKATVVTGSHASSTPGGKETTSATQSSVPSSTERNAYSMTSSVLSHS 74

QY 134 P-KAETTTKGP--ALTTKPEP-----TPTTPKEPASTTP-----KE 166
Db 75 PGSGSSTTQGDVTLAPATEPAGSAAATGQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134

QY 167 PTTT-----IKSAPTTKPEAPTT-----KSAPTTKPEAPTTTKEP-----A 206
Db 135 PAPGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTS 190

QY 207 PTPPKPEAPTTTKEPAPTTTTSAPTTKPEAPTT-----KKPAPTTKPEP----- 252
Db 191 PDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGV 247

QY 253 --APTTPKEPTPTTP-----KEPAPTTKPEP-----APTTPKPEAPTPAK----- 289
Db 248 TSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGV 307

QY 290 KPAPTTKPEAPTPK-----EPAPTTTKEPSPTTTPKEPAPTT--TKSAPTTTKEPAPTT 342
Db 308 TSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGV 367

QY 343 TKSAPTTKPEPSPTTTPKPEP-----APTTPKPEAPTTK-----KPAPTTKPEAPTTK 391
Db 368 T-SAPDT--RPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPA 424

QY 392 ---EPAPTTTKEPAPTTKPEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 447
Db 425 HGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPA 484

QY 448 ---ELAPTTPEPTPTTP-----EPAP--TTPKA-----AAPNTKPEAPTTK 487
Db 485 HGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGSTAPPA 544

QY 488 ---EPAPTTTKEPAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 534
Db 545 HGVTSAPDTRAPAGSTAP-----PAHGVTSAPDTRAPAGSTAPPAHGVTSAPDTRAPAGST 600

QY 535 APTTTKEPTSTSDKAP--TTPKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 588
Db 601 AP-PAHGVTSAPDTRAPAGSTAPPAHGVTS-----APDTRAPAGSTAPPAHGVTSAPDTR 655

QY 589 EPAPTTT-----KKPAPKELAPTTTKEPTTSTSDKAP--TTPK-----ETAPTTT 632
Db 656 APGSTAPPAHGVTSAPDTRAPAGSTAP-PAHGVTSAPDTRAPAGSTAPPAHGVTSAPDTR 714

QY 633 KEPAPTT-----KKPAPTTTPEPTTSEVSTPTTKEPTT-----IHKSPD 675

Db 715 PAPGSTAPPAHGVTSAPDTRPAPGS--TAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPD 772
QY 676 ESTPELSAEPPTKALENSKPEGVTTTKTAA-----TKPEMTTAKDK---TTERD 724
Db 773 TRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPD 832
QY 725 LRTTPETT-----TAAPKMKETATTTTEKTESKITATTTQVSTTTTQDTP--FKIT 775
Db 833 TRPAPGSTAPPAHGVTSAP-----DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGV 888
QY 776 TLKTTTTLAPKVTTTKTITTT--EIMNKPEETAAPKADKDRATNSKATTPKPKPTKAPKPT 833
Db 889 SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAP--GSTAPPAHG 946
QY 834 STKKPKTMRVRKPKTTPTRKMTSTMPENP-----TSRIAEAMLQ--TTRP---NQTN 885
Db 947 VTSAPDTRP--APGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 1003
QY 886 SKLVEYNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPVPMOGIINPMLSDETN 945
Db 1004 AHGVTSAPDTRPAPGSTAPPAH-----GVTSAPDNRPALGSTA---PPVHNVTIS 1049
QY 946 ICNGKPVGGLTTLRNGT 962
Db 1050 ASGSASGSASTLVHNGT 1066

RESULT 10
T25697
hypotheoretical protein F16F9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T25697
R:Fullon, B.
submitted to the EMBL Data Library, August 1996
A:Description: The sequence of C. elegans cosmid F16F9.
A:Reference number: Z20071
A:Accession: T25697
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1229 <FUL>
A:Cross-references: EMBL:U67956; PIDN:AB07691.1; GSPDB:GN00028; CESP:F16F9.2
A:Experimental source: strain Bristol N2; clone F16F9
C:Genetics:
A:Gene: CESP:F16F9.2
A:Map position: X
A:Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 12.8%; Score 659.5; DB 2; Length 1229;
Best Local Similarity 29.0%; Pred. No. 2.4e-21;
Matches 307; Conservative 78; Mismatches 339; Indels 333; Gaps 52;

QY 33 KVTT-PDSTTQ--HNKYSTSPKIITA-KPINRPSLPNSDT-----SKETSITVN 80
Db 119 KVTTSDASTNAPTCKDSTTPEITIGIVINSKSESVDMSSTTRFSTLSPTTELLS 178

QY 81 KETTVEKETTNNKOTSTDGKEKTTSAKETQSIKTSKADLAPTSKVLAKPPKAETTT 140
Db 179 PFTLVSTDSSTST-EQTSNDNTEIASPMETNTTTEATTTSEVPSVSTLAS---EDETIV 234

QY 141 KGPALTPKEPTTTPKPEASTTPKEPTTTPKESAPTTTKEPAPTTTTSAPTTTKEPAPT 200
Db 235 TATAEST-----TTVIAEVSTTTEPTTT-----AESTTKKSTT-----KAPA 272

QY 201 TTKEPAPTTKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 256
Db 273 TTEPTPTTTEE--VTTTEAETSTTTSSETSTEBK---PTTPLIDNKIAGPATGK---PET 324

QY 257 PKEP--TPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 314
Db 325 THFPVGTTFN--FDATETPTFVAKSDEKMTLSKTAATETTQQTTEVT-DGPEKETTKNV 381

[illegible]

```

RESULT 11
T11622
  extensin class 1 precursor - cowpea
  C:Species: Vigna unguiculata (cowpea)
  C:Date: 16-Jul-1999 #sequence_revise
  C:Accession: T11622; S54155
  R:Arsenijevic-Maksimovic, I.; Broughton, W.J.; Krause, A.
  Mol. Plant Microbe Interact. 10, 93-101, 1997
  A:Title: Rhizobia modulate root-hair-specific expression of extensin genes.
  A:Reference number: Z17301; NUID:97155574
  A:Accession: T11622
  A:Status: preliminary; translated from GB/EMBL/DBJ
  A:Molecule type: DNA
  A:Residues: 1-489 <R>
  A:Cross-references: EMBL:X91836; NID:g1015936; PIDN:CAA62943.1; PID:g101593
  A:Experimental source: sub-species Red calcoona
  R:Arsenijevic-Maksimovic, I.; Broughton, W.J.; Krause, A.
  Submitted to the EMBL Data Library, April 1995
  A:Description: A class of root-hair specific extensins involved in rhizobium

```

A:Reference number: S54155
A:Accession: S54155
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 326-489 <AR2>
A:Cross-references: EMBL:X86030; NID:g791149; PID:g791150
C:Genetics:
A:Gene: Ext26G
C:Superfamily: hydroxyproline-rich glycoprotein
C:Keywords: glycoprotein; hydroxyproline
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-489/Product: extensin class 1 #status predicted <WAT>

Query Match 12.3%; Score 633; DB 2; Length 489;
Best Local Similarity 31.2%; Pred. No. 1.4e-20;
Matches 149; Conservative 33; Mismatches 252; Indels 44; Gaps 7;

[illegible]

```

RESULT 12
138346
elastic titin - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_ref
C:Accession: I38346
R:Labalet, S.; Kolmerer, B.
Science 270, 293-296, 1995
A:Title: Titins: giant proteins
A:Reference number: A57430; MUID
A:Accession: I38346
A:Status: preliminary; translated
A:Residues: 1-7962 <RES>
A:Cross-references: EMBL:X90569;
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:I27867;
A:Map position: 2q31-q31

```

Query Match	Score	DB 2	Length
12.38	632.5	DB 2	7962


```
QY 503 KETAPTPKGTARTTLKEPAPT--TPKKPAPKELAPTTTKEPTSTTSXKPAPTTPKGTAP 560
Db 481 TSPVPVQV---PPVQKPPPTYSPPVKPPPIQKPEPT---PTYSPPIKPPVKKPPTPT 532
QY 561 TTPKEPAPTTPKEPAPT-TPKGTARTTLKEPAPT--TPKKPAPKELAPTTTGTSTTSD 617
Db 533 SPPIKP-PPVHKPPPTTYSPPPIKPPPIHKKPPTTYSPPPIKPPVHKPPTTYSPPPIKPPP 591
QY 618 KPAPTTPKETAPTTPKEPAPTTPKPKPAPT-TPETPPPTTSEVSTPTTK--EPTTIHKSP 674
Db 592 VHKPPPTTYSPIKP---PPVHKPPTTYSPIPKPPVHKPPTTYSPIPKPPVHKPP 647
QY 675 DESTPELSAEPKPALENSKPEGVPTTKTPAAKPEMTTAKDKTERDLRTPETTTA 734
Db 648 ---TPTYSPPPIKPPVQKPPPTTYSPPVKKPPVQLPP-TPTYSPPVKKPPVQVPPPTYS 703
QY 735 APKMTKETATTTTEKTESKITATTTQVSTTTQDTP 771
Db 704 PPVKKPPVQVPPPTTYSPIKPPVQVPPPTTTPSP 740

RESULT 14
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favella, A.; Vaudin, M.
submitted to The EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
3504/1

Query Match 12.2%; Score 629; DB 2; Length 3507;
Best Local Similarity 24.1%; Pred. No. 1.2e-19;
Matches 252; Conservative 158; Mismatches 416; Indels 218; Gaps 45;

QY 10 KKKPTKPPVDEAGSLDN-GDFKVT---PDTSTQHNVKYSTSPKITTA-----K 57
Db 1942 KQPNREKIEIDENSSNSGQEKPTTKGIVSSTSATSSSTAEAPHVTTTSSSTTSNK 2001
QY 58 PINPRESLPN---SDTSKETSIVNKEIV-ETKETTITNQTSDGKEKTSIAKETQS 113
Db 2002 DMTSSK-PENVMTSESPEVSTSSKSTAGETTVSSTPSSSSSEAPLTSAPATTEV 2060
QY 114 IKTSAKDIAP---TSVKLAPTKPAETTKGPAITLTPKEPTTPKPEASTTPKEPP 169
Db 2061 ITSSSVKSTTPKESSSEITVLSSKSPEVTSVSSKSSSTPS-TTSQSVTSTVPTSKS 2119
QY 170 TTIKS-APTTPKEPATTTKSAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTKS 228
Db 2120 TVLSSAPVTSPTSEVHT-SSETKPSLSASSTGDTNSTTPTSSSLASVKSTSAPEGTS 2178
QY 229 APTTPKEPATTTPKPAITPK-EPAPTTPKEPTTPKEPATTTPKEPAPTTPKEPAPT 286
Db 2179 ASVAPVKLSLSDVQSPSTKTFDATESVTQASETSSTSVKSTSEPESHVTKLSITS 2238
QY 287 APKKPAPTTPKPAITTPKEPATTTPKPSPTP--KEPAPTTPKSAPTTPKEPATTTK 344
Db 2239 NFSSSVPTVSPKSTPIVPE-----STEQPTSTPSCQSLTPMNSNSELVITSEPHVLSS 2293
QY 345 SAP-----TTPKEPSPTTKPAPTTPKPAITTPKPAITTPKPAITTPKPAITTP 397
```

```
Db 2294 LSPDVQSSTPNLSSSESTVE---TPKTSSEVSLNSEPSTTEAPTTLSLDILSTTN 2349
QY 398 -----TKKPAATAPK-----EPAPTTPKETAPTTPKKLTP 427
Db 2350 NLSQSTVSTEDKSETSSSENSEKPT-SAPELVTSVTHVASSSPDPVPTES--SEPDLTG 2406
QY 428 TTPKEKLAPTTPKEPAPTTPPEELAPTTPPEPTPTTPEP-----APTTPKAAAPN 476
Db 2407 SSTENIPEASSKGTISTTPTPDTTASEEPTKSTSPDLSTTNSVLSSESTPESSE-- 2464
QY 477 TPKEPAPT-----PKPEAPT-----TPKEPAP-----TTPKETAPTTP 510
Db 2465 --KSPVSSSTEGISVSTSTEFKSVPESTISSVLEEDLTKTTPSVILEETTTASETSEPLT 2522
QY 511 KGTAPTTLKEPAPTTPKPAKELAPTTTKEPTSTTSDKPA-----PT----- 553
Db 2523 EDSLTVSVRIHELTTSSENVPKESSTTTSSSESKSPQEPAGILTSTVVVPTSSVSLITA 2582
QY 554 -----TP--KGTAP--TTPK-----EPAPTTPKEPAPTTPKGTAPTTLKEPAPT 593
Db 2583 SEIEAITSNTPPKQGRPTTTPSKSLVKSTTSSTVTSSEPSSESTKRTTVSTTVSTTPT 2642
QY 594 TPKKPAPKELAPTTTKGPT-STTSDKPAPTTPKETAPTTPKEPAPTTPKPAITTPPETTP 652
Db 2643 EETTTSSELILTAAPSKPTESTTESSEAPTTPAKTSETKPSNVSSSTSRKSTENVEITSTQ 2702
QY 653 PTTSEVSTPTTKETPTTIHKSPDESTP--ELSAEPTPKALE--NSPKPEGVPTTKTPAAT 708
Db 2703 SCSLESTMSSTS-----SEPETNAPAVTVSSSEASTTLEENSSSTSSP----TSSEASV 2752
QY 709 K-----PEMTTAKDKTTERDLRTPETTTAAPKMTKETATTTTEKTESKITATTT--- 759
Db 2753 KLSSLPESITSEAVTVSSR----APAEITMSSSEHIREISTVSSESEPEIPLSTTVSPN 2808
QY 760 -----QVTSTTQDTP-FTITT-----LKTTLTLPKVVTTTKKTTITTEIMNKP 802
Db 2809 VVTASSIPSEEPILSSVTSSTPRVRLITGTPDDLIVSVTVPSHGNNRRONITASSV---P 2865
QY 803 EETAKPKDRATNSKATTPKPKPTKAPKAPTSTKKPKTTPRVRKPK---TTPTPKMTST 859
Db 2866 SNSTSPIILPSES-LTTPQPPPTTTTAKPATTSGRGPSPSIQPPAEEMETTPAP----- 2918
QY 860 MPELN-----PTSRIAEAMLQTTT 878
Db 2919 PPSNGGYGEETNQEEQVSTTTT 2942

RESULT 15
I51618
nucleolar phosphoprotein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
C:Accession: I51618; S57757
R:Cafrins, C.; McStay, B.
J. Cell Sci. 108, 3339-3347, 1995
A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp
A:Reference number: I51618; MUID:96019267
A:Accession: I51618
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-990 <CAI>
A:Cross-references: EMBL:X88927; NID:g895920; PIDN:CAA61368.1; PID:g895921
C:Genetics:
A:Gene: xNopp180
C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

Query Match 12.2%; Score 626.5; DB 2; Length 990;
Best Local Similarity 26.2%; Pred. No. 4.9e-20;
Matches 253; Conservative 132; Mismatches 393; Indels 187; Gaps 45;
```

Qy	3	DNKNRTKKPTKPPVVDGAGSLDNGDFKVYTTDDTSTQHNKVSTSPKTIITAKPIINPR	62
Db	59	DAKKRRPPANGLPKKKSAKESSEDSEEDPEPPAKKRAQ-----PAGGKPVVKAQPK	114
Qy	63	PSLPNPSDTSKETSLVWNKETTVE--TKETTTNK-----OTSDGKEKITSAKETQSIEK	116
Db	115	KAKSSEDSDESD--SEETKKPKAKRPAQTPKVAAYKTQTQKAKSSS-ESSSED	170
Qy	117	TSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKE-----PASTTPKETPTPTI	172
Db	171	EASKKQPVTKV---PPQAVVKAGLAGNNGKTADSSSESDSPAPKAKTAATTPPT-	225
Qy	173	KSAPTTPKBPATTTKSAPTTKBPATTTTKBPATTTTKBPATTT--TKBPATTT--T	226
Db	226	--RPATAAKFOAKTAGKSSSEDSDEEQTKAKSKPDIYVAPPTSVYSKKKT	283
Qy	227	KSAPTTPKBPAPT-----TPKKPAPTTPKBPAPTT---KEPTTTTPKBPAPTT	272
Db	284	LSOPGTKAKPESDSDSDSEEBQPAKAKIIPVAKAAASAPKAPLAKAETSDSESDSS	343
Qy	273	KEPAPTTPKBPAPTAPKKPAPTTPKBPAPTT---KEPAP-----TT	311
Db	344	EDEKSSVGLGVAAKPK-APAAP--DAKSTPVAAKKSAPAKKASSSDSDSSNEETT	400
Qy	312	TKBPSPPTPKBPAPTTTKSAPTTTKBPAPTTTKSAPTTPKPSPTTTKBPAPTTKPEP-	370
Db	401	TKPAAKTTPAKSAATPTSKTPTNGK--ATPTSKTPAKPGCTPKTSTAKKSDSSSDSS	457
Qy	371	--PTTPKKPAPTTKBPAPT--TPKBPAPTTTKKPAPTAPKBPAP-PTTPKETA-----	418
Db	458	SDEETTTKPAKTTTPAKSAATPTSKTP--TNSKATPTSKTTPAKPGCTKSAKKDSSS	514
Qy	419	-----PTTPKKLTTPTEKLIAPTTPKBPAPTTPEELAPTTPEEPTTTPEEPAPTT-PA	472
Db	515	SDSSDSSDSEKKTPA--KRAAKTTAPKA-----AKTTAPKA-----AKTTPAKP	558
Qy	473	AAPNTKBPAPT---TPKBPAPTTPKBPAPTTPKETAPTT-KGTAPTTLKEPAPTTPK	527
Db	559	AAKSTPGKQVPTKESSDSDSEDEKKSAKPAVKITFGKATS-----KPVVAS	611
Qy	528	KPAKELAPTTTKEPTSTTSDDKPAPTTPKGTAPTTPKBPAPT--PKBPAPTTPKGTAPT	585
Db	612	KPVPAK-----KASSSDSDSEEBETTTTKPLTKLSPAVKTLPPKKAESSDSDSSDSD	665
Qy	586	TLKEPAPTTPKKPAKELAPTTTKG-----TSTTSDKPAPTTPKETAPT--TPKEP	635
Db	666	SEKK--TKPAKPAKSAIPVNTKAPQONKASKASCDSDSDSEEGKSKQPTGKSPAAK	722
Qy	636	APTPPK-PAPTTPTPTPTSEVSTPTTTKE-PTTIHKSPDESTPELSAETPKALENS	693
Db	723	ATAPPKNPVAVNKDPSSSSDSDSGDDEKQKPKQAAAANKVQKAKAAPTTPKKAASS	782
Qy	694	PKBPGVPTTKPAATKPEMTTTAKDKTTERDLRTPTTETTTAAPKMTKETATUTKTESK	753
Db	783	SSE---DSSDEDSVAKKNTNTAVKSPV---TTPKAVPAKKSSSESSSEDEKQGG	835
Qy	754	ITATTTQVSTTTQDTPPKITLTKTTTLAPKVTTTK--KTITTEIMNKPEETA-KPKD	810
Db	836	KNTSTTKIANST-----PRAAAACESSSSDECKANGTSGKRRK	877
Qy	811	RAYNS--KATTPKPKPTKAPKKPTSKKPKTPMPVRKPKTPTP-RKMTSTMPELNPT	866
Db	878	ESTGNAECEAVTEP-----NKKLUKASPTFPVKNKKELKNTPPRFFVEEDIEINP-	928
Qy	867	SRIAE 871	
Db	929	-RWAD 932	


```
QY 871 EAMLTQTTTREPNOTPNSKLVNPKSADAGAGETPHMLLRPHVFMVETPDMDYLP 927
SQ SEQUENCE 1367 AA; 874 N-LINKED (GLCNAC...) (POTENTIAL),
      1367 AA; 91COOE2DBD61AA9D-CRC64;

Query Match 14.8%; Score 763; DB 1; Length 1367;
Best Local Similarity 28.5%; Pred. No. 1.1e-23;
Matches 286; Conservative 98; Mismatches 463; Indels 154; Gaps 43;

QY 35 TTPDTSTTOHNKVSPTKITTAKPINP-----RPSLPPNSDTSKETSITVNKETTVE 86
DB 236 STSESSTTSSTSESSTSSSTTAPATPTTSCTEKPTPTTSCTEKPTPHHDTPC 295
QY 87 TKETTTNKOTSDGKEKTSKQSEKTSKAKDLAPTSKVLAKPTPKAETTT--TKGPA 144
DB 296 TKKKTTSK-TCT--KKTTPVTPS-SSTTESSAPV-----PTFSSSTESSAPV 344
QY 145 LTPKE-----PTPTTKEPASTTTPKEPTTTPKAP-----TTPKEPATTTKSAPTTP 194
DB 345 TSSTESSAPVPTPSSSTESSAPVTSSTESSAPVTSSTESSAPVPTPSSSITE 404
QY 195 KEPAPTTTKEPAPTTTKEPAPTTTTSAPTTPKEPAPTTTTPKAPTTTKEPAP 254
DB 405 SSSAPVTS-----TTESSAPVTS--STTESSAPVT--SSTESSAPVTSSTES 453
QY 255 TTPKEPTT--TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 307
DB 454 SSAPVPTPSSSTESSAPVT--SSTESSAP-VPTPSSSTESSAPVPTPSSSTESS 509
QY 308 APTTTPKEPTTTPKEPAPTTTTSAPTTPKEPAP-----TTKAPTTPKPSPTTTPKEPAP 363
DB 510 APVPTPS-SSTESSAPAPTSSSTESSAPVTSSTESSAPVTSSTESSAPVPTPSSSTESS 568
QY 364 ---TTPKEPAPT-TPKPAPTTPKEPAPTTPKEPAPTTPKAPTTTTPKAPTTTKEPAP 418
DB 569 VTSSTESSAPVPTPSSSTESSAPVT--PSSSTESSAPAPTSSSTESSAPVPTPSSSTESS 624
QY 419 PTPPKLTPTPKLPAPTTPKEPAPTTPPELAPT-TPEEPTTPTPKEPAP-----TTPKA 472
DB 625 PVT-----SSTESSAP-VPTPSSSTESSAPVPTPSSSTESSAPVPTPSSSTESS 679
QY 473 AAPNTKEPAPTTPKEPAPTTPKEPAPTTPKAPT-TPKAPTTPKAPTTTTPKAPTTTTPKAP 531
DB 680 SAVPT-----SSTESSAPVT--SSTESSAPVPTPSSSTESSAPVPTPSSSTESS 731
QY 532 KELAPTTTKEPTSTSDKAPPTPKGTAPTTPKEPAP---TTPKEPAPT-TPKGTAPTIL 587
DB 732 SSSAPVPTPS-SSTESSAPVTSSTESSAPVPTPSSSTESSAPVPTPSSSTESS 790
QY 588 KEPAPTTPKKAPKELAPTTTGTPTSTSDKAPTTTKEPAPTTPKAPTTPKEPAP-----TTPKK 642
DB 791 SAVPTPSSSTESSAPVPTPSSSNISSAPSSSTESSAPVPTPSSSTESS 850
QY 643 PAPTTPETPPTTSEVSTPT-----TKETTHSPDESTPELSAETPKALENSKPEP 697
DB 851 SAVSSSTESSAPVPTPSSSNISSAPSSI--PFSSTESFSTGT-TVTPSSSKVP 906
QY 698 GVPTTKPAATKEMMTTAKDKTTERDLRTPTTTTAAAPKMKETATTTTEKTESKITAT 757
DB 907 GSQTETSVSSTETTIVPTTKTTTSVTTPTTTTTCSTGTNSAGETSGCSPKVTVT 966
QY 758 --TQVTSSTTQDTPFKLTLTKTLTTLAPKVT--TKKTTT-EIMNKEPETAAPK 811
DB 967 VPTTTTTSVTSSTTTTTCSTGTNSAGETSGCSPKTIITTVPCSSPSETA-----1021
QY 812 ATNSKATTPKPKP-----TKAPKPTSTKPKMTPVRKPKTTPTPR 854
DB 1022 ---SESTTSPPTPVTTVVSTTVVTEYSTKPGGEITTTFTVKNIPTLYLTIAPT-1077
QY 855 KMTSTPELNPSTRIAEAMLQTTTRNQTPNSKLVNPKSADAGAGETPHMLLRPHV 914
DB 1078 -SVTTVTFNFTPTT-ITTVTCSTGT-----NSAGETSGCSPKVTITVTPCST 1122
QY 915 FMEVPTPDMDYLPVNPNOGIIINPMLSDETNICNGKPVVDGLTT 957
```

Db 1123 GRGVTTEATLTVTAVTTVTTESSTGTSNSA-GKTTTGYTT 1164

RESULT 4

MUC1_HUMAN

AC AC P15941; P15942; P15931; P17626; Q14128; Q16442; Q16437; Q9Y4J2; PRT; 1255 AA.

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM) (PEMT)

DE (EPITHELIAL) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)

DE (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H23AG) (PEANUT-

DE REACTIVE URINARY MUCIN) (PUM) (BREAST CARCINOMA-ASSOCIATED ANTIGEN

DE DF3).

GN MUC1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Pancreas;

RX MEDLINE=90368716; PubMed=2394722;

RA Ian M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;

RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";

RJ J. Biol. Chem. 265:15294-15299(1990).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=90207794; PubMed=2318825;

RA Lightenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;

RT "Epistatin, a carcinoma-associated mucin, is generated by a

RT polymorphic gene encoding splice variants with alternative amino

RT termin".

RJ J. Biol. Chem. 265:5573-5578(1990).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Breast carcinoma;

RX MEDLINE=90368715; PubMed=1697589;

RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,

RA Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;

RT "Molecular cloning and expression of the human tumor-associated

RT polymorphic epithelial mucin.";

RJ J. Biol. Chem. 265:15286-15293(1990).

RN [4]

RP SEQUENCE FROM N.A.

RX MEDLINE=91097524; PubMed=2268309;

RA Lancaster C.A., Peat N., Duhig T., Wilson D.,

RA Taylor-Papadimitriou J., Gendler S.J.;

RT "Structure and expression of the human polymorphic epithelial mucin

RT gene: an expressed VNTR unit.";

RJ Biochem. Biophys. Res. Commun. 173:1019-1029(1990).

RN [5]

RP SEQUENCE FROM N.A.

RC TISSUE=Breast carcinoma;

RX MEDLINE=90276413; PubMed=2351132;

RA Wreschener D.H., Hareuveni M., Tsarfaty I., Smorodinsky N.,

RA Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,

RA Keydar I.;

RT "Human epithelial tumor antigen cDNA sequences. Differential splicing

RT may generate multiple protein forms.";

RJ Eur. J. Biochem. 189:463-473(1990).

RN [6]

RP SEQUENCE FROM N.A.

RC TISSUE=Breast carcinoma;

RX MEDLINE=90276414; PubMed=2112460;

RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,

RA Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschener D.H.;

RT "A transcribed gene, containing a variable number of tandem repeats,

RT codes for a human epithelial tumor antigen. cDNA cloning, expression

RT of the transfected gene and over-expression in breast cancer

RT tissue.";

RJ Eur. J. Biochem. 189:475-486(1990).

RN [7]

RP SEQUENCE FROM N.A.

RC TISSUE=Breast carcinoma;

RX MEDLINE=91033045; PubMed=1688329;

RA Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,

RA Jeitsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschener D.H.;

RT "Isolation and characterization of an expressed hypervariable gene

RT coding for a breast-cancer-associated antigen.";

RJ Gene 93:313-318(1990).

RN [8]

RP PARTIAL SEQUENCE FROM N.A.

RX MEDLINE=88330762; PubMed=34117635;

RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,

RA Burchell J.;

RT "A highly immunogenic region of a human polymorphic epithelial mucin

RT expressed by carcinomas is made up of tandem repeats.";

RJ J. Biol. Chem. 263:12820-12823(1988).

RN [9]

RP SEQUENCE OF 1-169 FROM N.A.

RX MEDLINE=90088473; PubMed=2597151;

RA Abe M., Siddiqui J., Kufe D.;

RT "Sequence analysis of the 5' region of the human DF3 breast

RT carcinoma-associated antigen gene.";

RJ Biochem. Biophys. Res. Commun. 165:644-649(1989).

RN [10]

RP SEQUENCE OF 1-109 FROM N.A.

RC TISSUE=Thyroid;

RX MEDLINE=96183746; PubMed=8608966;

RA Weiss M., Baruch A., Keydar I., Wreschener D.H.;

RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse

RT transcriptase polymerase chain reaction of the MUC1 gene.";

RJ Int. J. Cancer 66:55-59(1996).

RN [11]

RP SEQUENCE OF 1-89 FROM N.A.

RC TISSUE=Lung;

RX MEDLINE=96181716; PubMed=8604237;

RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,

RA Lee L.N., Luh K.T., Wu C.W.;

RT "Mucin mRNA expression in lung adenocarcinoma cell lines and

RT tissues.";

RJ Oncology 53:118-126(1996).

RN [12]

RP SEQUENCE OF 1-46 FROM N.A.

RC TISSUE=Breast carcinoma;

RA Bulwella L., Liu Q., Luqmani Y.A., Gomm J.J., Coombes R.C.;

RT Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN

CC CYTOSKELETON.

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM

CC IS ALSO PRODUCED.

CC -!- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE

CC SPLICING.

CC -!- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL

CC TUMORS, SUCH AS BREAST CANCER.

CC -!- PTM: HIGHLY GLYCOSYLATED (N-AND O-LINKED CARBOHYDRATES AND SIALIC

CC ACID).

CC -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT

CC VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE

CC MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.

CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; J05582; AAA60019.1; -

CC EMBL; M32738; AAA35804.1; -

CC EMBL; M32739; AAA35806.1; -

CC EMBL; J05581; AAA59876.1; -

protein.";
RL Biotechnol. Prog. 6:171-177(1990).
[2]
RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.
RX MEDLINE=83135732; PubMed=6298211;
RA Walte J.H.;
RT "Evidence for a repeating 3,4-dihydroxyphenylalanine- and
RT hydroxyproline-containing decapeptide in the adhesive protein of the
RT mussel, Mytilus edulis L.";
RL J. Biol. Chem. 258:2911-2915(1983).
CC -!- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
CC -!- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
CC -!- PTM: THE DECAPEPTIDE A-K-P-S-Y-P-T-Y-K IS POST-TRANSLATIONALLY
CC MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE
CC HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE
CC (DOPA) DERIVED FROM TYROSINE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X54422; CAA38294.1;
DR InterPro; IPR002964; Adhesive_plaq.
DR PRINTS; IPR002965; P_rich_extensn.
DR PRINTS; PR01216; ADHESIVEI.
DR PRINTS; PR01217; PRICHEXTENS.
KW Repeat; Hydroxylation.
FT NON_TER 1
FT DOMAIN 67 870
SQ SEQUENCE 875 AA; 100412 MW; 6EA85312748CAACE CRC64;
P-P-[ST].
TANDEM REPEATS OF Y-K-[PA]-K-[LP]-[ST]-[Y]
Query Match 10.8%; Score 556.5; DB 1; Length 875;
Best Local Similarity 28.7%; Pred. No. 8.9e-16;
Matches 272; Conservative 106; Mismatches 375; Indels 195; Gaps 56;
OY 42 TOHNNKVSPTLTTAKPNP-----RP--SLPPNSDTSKETSLLTVKETTIVET 87
DB 1 TKHEPVYKPVKTSYSAKPKPTQYQPLKKVDYRPTKSYPPYVG-SKTYLPLAKKLSSYK 59
OY 88 KETTTNNKQTSVG--KEKTT---SAKETQSIKETSADLAPTSKVLAKPTPKAETTTKG 142
DB 60 PIKTYNAKTNYPVYKPKMTYPPYKPKPSYPPYKSKYKPKIITYPPYKAKPSY-- 117
OY 143 PALITPKETPTTPEK-----PASITPKETPTTTSAPTTP-----KEPAPTTKSAPT 193
DB 118 PSSYKPKRTYPPYKPKLTYPPYKPKPSYPPYKPKPSYPPYKPKPSYPPYKPKPSY 177
OY 194 KPEAPTPTTKEPA--PTTPKPEA-----PTTTTKEPA--PTTTKSAPTTPK--EPAPT 240
DB 178 P-----PTTKAKPSYPPYKAKPSYPPYKAKPTTKAKPTTKAKPSYPPYKAKPT 233
OY 241 PKKPA--PTTPKPEA--PTTPKPEA--PTTPKPEA--PTTPKPEA--PTAPK 289
DB 234 KAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPTTKAKPTTKAKPSYPPYKAKPSY 293
OY 290 KPA--PTTPKPEA--PTTPKPEA--PTTPKPEA--PTTPKPEA--PTTPKPEA--PTTP 339
DB 294 KPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPT 353
OY 340 PTTTKSAPTTPKPSPTTTPKPA--PTTPKPEA--PTTPKPEA--PTTPKPEA--PTTPKPEA 395
DB 354 PSTYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSY 409

OY 396 TTTTKKPAPTAPKPAETTPKETAPTTPKLTPTTPEKLAETTPKETAPTTPPEELAPTPE 455
DB 410 PPTYKAKPTYKAKP-----TYPST-YKAKPSY-----PSYKAKPSYPPYKAKPTTKA 457
OY 456 EPTPTTPEEPAPTTPKAAAPNTPKEAPTTPKEAPTTPKE-----PARTTPKETAPTTP 509
DB 458 KPT-----YSTYKA-----KPSYASAKPSYPPYKSSYSSYKPKKTYPPPTTK 506
OY 510 PKGTAPTTLKEAPTTPKPAKELAPTTPKPTSTSDKPAETTPKGTAPTTP-----PKE 565
DB 507 PKLTYKPTYK-PKPSYPPSYKPKTTPYPTTK-----PKISYPPYKAKPSY 551
OY 566 PARTTPKEAPTTPKGT-----APTTLKEPA--PTTPKKPA--PKELAPTTPKGT--PTSTSD 617
DB 552 PATYKAKPSYPPYKAKPSYPPYKAKPSYKAKPTVPSTYKAKPSYPPY 611
OY 618 KPAPTTPKETAPTTPKPA--PTTPKKPA--PTT-----PETPPTTSEVSTPTTKE---- 566
DB 612 KAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSY 667
OY 667 PTTIHKSPDESTPELSAETPKALENSPK-----EPGVPTT--KTPA-----ATKPEMTT 715
DB 668 PPT-YKAKPSYPPYKAKPTTKAKPTNPSTYKAKPSYPPYKAKPSYPPYKAKPSYPT 726
OY 716 AKDKTTERDLRTPETTTAAPTETATTTETTESKITATTTQVTTSTTQDTPPKIT 775
DB 727 YKAKPTTKAKPTVPSTYKAKP--TYKAKPTYPYKAK-----PSYPTTKPKPSYPP 777
OY 776 TLKTTTLAPKVTITTKTITTTIMNKPEETAKPKORATNSKATTPKQKPTKAPK--P 832
DB 778 TYKSKSIYPPSYKPKRTYPTT---YKPLUTYPTTK-----PKSPYSPYKPKRTYPT 826
OY 833 TSTKKPKTMRVVRKPKTTPTP-----RKMTSTMPELNPTSR 868
DB 827 STYKLPSPYPPYKSKTSYPPYKSKISYPPSYKAKTSYPPAYKPTNR 874
RESULT 6
ZAN_HUMAN
ID ZAN_HUMAN STANDARD; PRT; 2700 AA.
AC Q9Y493; O00218;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ZONADHESIN (FRAGMENT).
GN ZAN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-2379 FROM N.A.
RX MEDLINE=99018118; PubMed=9799793;
RA Glockner G., Scherer S., Schattevov R., Borlight A., Weber J.,
RA Tsui L.C., Rosenthal A.;
RT "Large-scale sequencing of two regions in human chromosome 7q22:
RT analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci
RT reveals 17 genes.";
RL Genome Res. 8:1060-1073(1998).
RN [2]
RP SEQUENCE OF 2338-2700 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=97271566; PubMed=9126492;
RA Gao Z., Harumi T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RT zonadhesin gene (ZAN).";
RL Genomics 41:119-122(1997).
CC -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE

CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).

CC -|- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.

CC -|- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE

CC ZONA PELLUCIDA.

CC -|- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,

CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF

CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHUS.

CC -|- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT

CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).

CC -|- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.

CC -|- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC The European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AF053356; AAC78790.1; -.

CC EMBL; U83191; AAC51208.1; -.

CC MIM; 602372; -.

CC InterPro; IPR000561; EGF-like.

CC InterPro; IPR000998; MAM.

CC InterPro; IPR002965; P_rich_extensn.

CC InterPro; IPR002919; Tila.

CC InterPro; IPR003328; Tila.

CC InterPro; IPR001007; VWFC.

CC InterPro; IPR001846; VWFD.

CC Pfam; PF00629; MAM; 3.

CC Pfam; PF01826; TIL; 4.

CC Pfam; PF02345; Tila; 4.

CC Pfam; PF00094; vwd; 4.

CC PRINTS; PR01217; PRICHEXTENS.

CC SMART; SM00137; MAM; 2.

CC SMART; SM00214; VWC; 1.

CC SMART; SM00011; VWC def; 3.

CC SMART; SM00216; VWD; 3.

CC PROSITE; PS01186; EGF_2; 3.

CC PROSITE; PS00740; MAM_1; 1.

CC PROSITE; PS00600; MAM_2; 4.

CC Glycoprotein; Transmembrane; Cell adhesion; Repeat.

CC NON_TER 1

CC DOMAIN 1 109

CC DOMAIN 112 136

CC DOMAIN 161 326

CC DOMAIN 322 446

CC DOMAIN 483 951

CC 953 1065

CC DOMAIN 1066 1454

CC DOMAIN 1455 1861

CC DOMAIN 1862 2292

CC DOMAIN 2293 2684

CC ?

CC EGF-LIKE.

CC CARBOHYD 74 74

CC CARBOHYD 403 403

CC CARBOHYD 1023 1023

CC CARBOHYD 1099 1099

CC CARBOHYD 1618 1618

CC CARBOHYD 1737 1737

CC CARBOHYD 1832 1832

CC CARBOHYD 1878 1878

CC CARBOHYD 2136 2136

CC CARBOHYD 2505 2505

CC CARBOHYD 2374 2379

CC CONFLICT 2700 2700

CC NON_TER 2700

CC SEQUENCE 2700 AA; 293013 MW; 80E60C0B12277B1 CRC64;

Query Match 10.8%; Score 555.5; DB 1; Length 2700;

Best Local Similarity 32.6%; Pred. No. 2.3e-15;

Matches 199; Conservative 64; Mismatches 240; Indels 107; Gaps 32;	
QY 58 PINRPSLPNDSKTSKSLVKNKQETVETKTTTNNKQSTGDKERTSAKETSIEKT 117	117
Db 445 PVKVLPELPVPVSS-----TGPSETGLTENPILSTK-----KPTVSIEKP 487	487
QY 118 SAKDLAPTSKVLAKPTPKAETTTGALTTTKEPTTPPKEPASTTPKEPT-----PTTI 172	172
Db 488 SVTTEKPT-----VPKEKPTIPTKPTISTEK---PTIPSEKPNMPSEKPIPESEKPTIL 539	539
QY 173 KSAPITPKEPAPTTTKSAPITPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKSAPT 231	231
Db 540 TEKPTIPSE-KPTIPSEKPTISTEKPTVPTTE-PTTTEETTTTMEEPVPTKPSIPT 596	596
QY 232 TPKEPAPTTPKKAPTTKEPAPTTPKKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTPAKKP 291	291
Db 597 --EKSIPEK---PTISMEETIISTEKPTICPEKPTIPTK---PTIPTKSTISPEK- 647	647
QY 292 APTTPKEPAPTTPKKAPTTTPKEPAPTTTPKEPAPTTTPKAPTTTKSAPTTTKEPAPTTTKSAPT 351	351
Db 648 -PTTPE-KPTIPTKPTISTEKPTIPTK-PTISPEKLTIPTEKLTIPT---EKPTIPT 701	701
QY 352 EPSPTTKEPAPTTTPKEPAPTTPKKPA-----PTTPKEPAPTTTPKEPAPTTTKKAPTAP 406	406
Db 702 EKPTISTEE--PTTPEETIISTEKPSIPMEKPTIPTETTSVEETIISTEKLTIPM-- 757	757
QY 407 KEAPPTPKETAPT-----TPKKLTPTTPEKLA-----PTTPEKAPTTPEELAPTTPEE 456	456
Db 758 EKPTISTEKPTIPTKPTISTEKLTIPT-EKLTIPTEKPTIPIETIISTEKLTIP--TIPE 814	814
QY 457 PPTTPKEPAPTTPKAAPNTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKGTAPT 516	516
Db 815 KETISPEKPTISTEK-----PTIPE-KPTIPE-----ETTISTEKLTIPT 855	855
QY 517 TLKEPAPTTPKKAPKELAPTTTKETSTT-----SDKAPPTPKGTAPTTPKEPAPTT 570	570
Db 856 --EKPTISPEKLTIPTEKPTISTEKPTIPIETIISTEKLTIPTKPTIPT---EKLALR 911	911
QY 571 PKEPAPTTPKGTAPTTLKAPAPTTPKKAPKELAPTTTKGTPTSTSDKAPTTPKET--A 628	628
Db 912 PPHPSPTA-TGLAALVNSPHAPSTPMTSV---ILGTTTTSRSTGMSCP-PNARYESCAC 966	966
QY 629 PTPPKAPPT 638	
Db 967 PASCKSPRPS 976	
RESULT 7	
EXTN_TOBAC STANDARD; PRT; 620 AA.	
AC P13983;	
DT 01-JAN-1990 (Rel. 13, Created)	
DT 01-JAN-1990 (Rel. 13, Last sequence update)	
DT 01-MAR-1992 (Rel. 21, Last annotation update)	
DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).	
GN HRGPNT3.	
OS Nicotiana tabacum (Common tobacco).	
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;	
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.	
OX NCBI_TaxID=4097;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=CV. XANTHI; TISSUE=Leaf;	
RX MEDLINE=90128263; PubMed=2612909;	
RA Keller B., Lamb C.J.;	
RT "Specific expression of a novel cell wall hydroxyproline-rich	
RT glycoprotein gene in lateral root initiation.;"	
RL Genes Dev. 3:1639-1646(1989).	
CC - - FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN	
CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE	
CC MAIN ROOT.	
CC - - SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.	

DR	Pfam: PF00038; filament; 1.	
DR	PROSITE; PS00226; IF; 1.	
KW	Intermediate filament; Coiled coil; Neurone; Phosphorylation;	
KW	Repeat.	
FT	DOMAIN 1 97 HEAD.	564
FT	DOMAIN 98 408 ROD.	623
FT	DOMAIN 409 1087 TAIL.	
FT	DOMAIN 436 517 GLU-RICH (ACIDIC).	
FT	DOMAIN 519 886 50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.	
FT	DOMAIN 887 1087 GLU/LYS-RICH.	
FT	DOMAIN 98 129 COIL 1A.	
FT	DOMAIN 130 141 LINKER 1.	
FT	DOMAIN 142 239 COIL 1B.	
FT	DOMAIN 240 261 LINKER 12.	
FT	DOMAIN 262 283 COIL 2A.	
FT	DOMAIN 284 287 LINKER 2.	
FT	DOMAIN 288 408 COIL 2B.	
FT	CONFLICT 133 133 K -> QA (IN REF. 2 AND 3).	
FT	CONFLICT 199 199 A -> AR (IN REF. 2 AND 3).	
FT	CONFLICT 281 281 S -> T (IN REF. 2 AND 3).	
FT	CONFLICT 492 492 L -> G (IN REF. 2 AND 3).	
FT	CONFLICT 551 551 P -> PREAKSP (IN REF. 3).	
FT	CONFLICT 689 712 MISSING (IN REF. 3).	
FT	CONFLICT 714 714 G -> A (IN REF. 3).	
FT	CONFLICT 814 814 V -> M (IN REF. 2 AND 3).	
FT	CONFLICT 843 843 T -> N (IN REF. 2 AND 3).	
FT	SEQUENCE 1087 AA; 116612 MW; 57BAC76A38EDICB9 CRC64;	
SQ		
Query Match	10.3%; Score 533; DB 1; Length 1087;	
Best Local Similarity	27.5%; Pred. No. 8.7e-15;	
Matches 216; Conservative 81; Mismatches 330; Indels 158; Gaps 38;		
QY	21 DEAGSLDNGDFKVT-----TPTSTQHINKVSTSPKTTAKPDPINPRSLPNSDTSKET 75	
DB	404 EECRIGFGSPSLTBGLPKIPSIST--HIKVKSEMIKWVE-----KSEKET 449	
QY	76 SLTVNKETIVEKTTTNTKQSTQCKEKTTSKETSQKADLAPTSKVLAKPTPK 135	
DB	450 VIVGQTEELRVTEGVEEDKEAQOGEAEAGEEKEEELAAATSPPAEEAASPEKE 509	
QY	136 AETTTKGPALTTPKEPTTPPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPPTPK 195	
DB	510 TKSRRVEEAKSPGEAKSPGEAKSPA-----EAKSPGEAKSPAEFPKSPA 564	
QY	196 EP-APTTPKEPAPTTTPKEPAPTTTKEP-----APTTPKSAPTTTPKEPA-----PTTPKKPA- 245	
DB	565 EPKSPAEAKSPA--EPKSPA--TVKSPGEAKSPSEAKS--PAKAKSPAEAKSPAEAKSPA 619	
QY	246 ---PTTPKEPA-----PTTPKEP-----TPTTPKEPAPTTTPKEPAPTTTPKEPA-----PTAPKK 290	
DB	620 AKSPAFAKSPAEAKSPATVKSPEAKSPSEAKSPA---EAKSPAFAKSPAEAKSPAEVKS 676	
QY	291 P-----APTTPKEPAPTTTPKEPAPTTTKSPPTTPKEPA-----PTTPKSAPTTTPKEPAPTT 342	
DB	677 PGEAKSPAEPKSPA--KSPA-----EVKSPAFAKSPAEVKSPEAKSPAEVKSPEAKSPA 729	
QY	343 TKSAPTTPEK--PSPTTPKEPA-----PTTPKEPAPTTTPKPKPAPTTTPKEPAPTTTPKEPAPTT 397	
DB	730 SPAVKSPEAKSPGEAKSPAEAKSPAFAKSPAEVKSPEAKSPAEVKSPEAKSPAEVKSPEAKSP 789	
QY	398 TKPAPTAPEKAPPTTPKETAPTTTPKLTTPTEKLAFTTPKEPAPTTPELAPTTPEEP 457	
DB	790 AKSPVKEDIKPPAEAKSPKAE--KSPVKEGAKPPKAEKALDVKSPAEQTPVQEEATVPTDI 848	
QY	458 TPTTPPEE--PAPTTKAAAPNTPKEPAPTT-----PKEPAPTT-----KEPAPTTTPK 505	
DB	849 RP---PEQVKSPAFAKSPAE--KEEAKTSKVPKAEVKSPEVKSPEVKSPEVKSPEVKSPEVKSPE 904	
QY	506 APTTPKGTAPTTTPKEPAPTTTPKPKAPKELAPTTTPKSTSD--KPAPTTTPKGTAPTTTP 563	
DB	905 TLTPKTEAKSKDEAPKAPKPKVEKKEKTPTEKPKDSTAFAKKEAGEKKKAVASEE 964	
QY	564 KEPATTPKPEAPPTPKGTAPTTTLKEPAPTTPKKPAKELAPTTTKGPTSTSDKPAPTT 623	
DB	965 ETAKLGVKEEA--KPKETETTTKTEADTKAKEPS-----KPTETEKP----- 1006	
QY	624 PKETAPTTKPEAPPTPKKPAKPTTPPTTSEVSTPTTTPKPTTIHKSPDSESTPELSA 683	
DB	1007 -----KKEMPAAPEK-----KDKKEKTTESRKPEEK----- 1034	
QY	684 EPTPKALENSPKPGVPTTKTTPAATKPEMTTAKDKTTERDLRTTETTTAAPKMTKETA 743	
DB	1035 ---PK-MAVKEDDKSLKEP--SKPKTEKAEKSSSTDQKESQPP-----KTTEDKA 1082	
QY	744 TTTEK 748	
DB	1083 TKGEK 1087	
RESULT 9		
TCNA_TRYCR	STANDARD; PRT; 1162 AA.	
ID TCNA_TRYCR		
AC P23253;		
DT 01-NOV-1991 (Rel. 20, Created)		
DT 01-NOV-1991 (Rel. 20, Last sequence update)		
DT 20-AUG-2001 (Rel. 40, Last annotation update)		
DE STALIDASE (EC 3.2.1.16) (NEURAMINIDASE) (NA) (MAJOR SURFACE ANTIGEN).		
GN TCNA.		
OS Trypanosoma cruzi.		
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.		
OX NCBI_TaxID=5693;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN-SILVIO X-10/4;		
RX MEDLINE=91277609; PubMed=1711561;		
RA Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,		
RA Prioli R.P.;		
RT "The Trypanosoma cruzi neuraminidase contains sequences similar to		
RT bacterial neuraminidases, YWTD repeats of the low density lipoprotein		
RT receptor, and type III modules of fibronectin.";		
RL J. Exp. Med. 174:179-191(1991).		
RN [2]		
RP SUBCELLULAR LOCATION.		
RX MEDLINE=91376547; PubMed=1896773;		
RA Prioli R.P., Mejia J.S., Aji T., Aikawa M., Pereira M.E.A.;		
RT "Trypanosoma cruzi: localization of neuraminidase on the surface of		
RT trypanomastigotes.";		
RL Trop. Med. Parasitol. 42:146-150(1991).		
CC -!- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN		
CC PARASITE INVASION OF CELLS.		
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC		
CC LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACYLNEURAMINYL		
CC RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACYLATED		
CC NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,		
CC GLYCOLIPIDS OR COLOMINIC ACID.		
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR		
CC (POSSIBLE).		
CC -!- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPMASTIGOTES, MINIMUM		
CC IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.		
CC -!- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT		
CC DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TCNA		
CC PROTEIN.		
CC -!- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.		
CC -!- SIMILARITY: CONTAINS 3 BNR REPEATS.		
CC -----		
CC This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC the European Bioinformatics Institute. There are no restrictions on its		
CC use by non-profit institutions as long as its content is in no way		
CC modified and this statement is not removed. Usage by and for commercial		
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC or send an email to license@isb-sib.ch).		
CC -----		
DR EMBL: M61732; AAA30255.1; --		
DR FIR: JH0557; JH0557.		

J. Mol. Evol. 43:348-356(1996).

1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.

1- SUBCELLULAR LOCATION: SECRETED.

1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.

1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.

1- PFM: PROLINES IN THE REPEAT ARE HYDROXYLATED (SINGLE OR DI-) AND ALSO TYROSINE (THUS PRODUCING DOPA = 3,4-DIHYDROXYPHENYLALANINE).

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; D63777; BAA09850.1; InterPro: IPR002964; Adhesive_plaq. InterPro: IPR002965; P-rich_extensn. PRINTS; PR01216; ADHESIVEI. PRINTS; PR01217; PRICHEXTENS. SIGNAL; Repeat; Hydroxylation. FT CHAIN 1 20 POTENTIAL. FT DOMAIN 21 872 ADHESIVE PLAQUE MATRIX PROTEIN. FT DOMAIN 21 41 NONREPEAT LINKER. FT DOMAIN 124 872 TANDEM REPEATS OF Y-K-[PS]-K-[IP]-[ST]-Y-P-[PST]-[ST]. FT DOMAIN 184 192 NONAPEPTIDE 1. FT DOMAIN 213 221 NONAPEPTIDE 2. FT SEQUENCE 872 AA; 101677 MW; 98CC70D/C75FF3C4 CRC64;

Query Match 10.1%; Score 518.5; DB 1; Length 872;
Best Local Similarity 29.2%; Pred. No. 2.7e-14;
Matches 276; Conservative 80; Mismatches 407; Indels 181; Gaps 54;

QY 19 VVDEAGSLDNGDFKVTTPDST--TOH-----NKVSTSPKITTAKPINRPSLPNS 69
Db 29 VVGSAYSGASAGAYK-TLPGSHPGYGVKHPVYKPMNKIPT-PYI--SKSYAPAYKPGY 84
QY 70 DTSKETSIVNKEETVETKEITTTNKOTSTDCKEKTSKAKTOSIEKTSKADLAPTSKVL 129
Db 85 YPTKRYQPTGYGKTNVPIYKPIAKLLSYKAIKITYPAYKAKTSVPPSYK-----HKIT 139
QY 130 AKPTPKAETTTKGPALTTPEKEPT-PTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTK 188
Db 140 YPTTK-----PKITYP--PTYKQPSYPPSYKPKTTPPTTK-----PKITYPPTTK 185
QY 189 SAPT-TPKEPAPT--TTKEPAPTTPEKEPAPTTTKEPAPTTKSAPTTPKEPAPTT--PKK 243
Db 186 RKPSYTPYKPKATYPPTKPKITYP---PTYKRPKPSYT-----PKYKTTTPPTTKPKI 236
QY 244 PAPTTPKEPAPT--PKEPTPTT--PKEPAPTTTKEPAPTTTPEKEPAPT-APKKAPAPTTP 296
Db 237 SYPSIYKPKASYVSSYKSKKTYPTTKYKISYPTTKPKPSYP---PTYKPKVTPPTTY 292
QY 297 KEPAPTTPKEPAPTTTKEPSPPTTPKEPAPTTTKEPAPT-----TTKEPAPTTKSAPTPK 351
Db 293 K-PKPSYPTTKPKITYPPT-YPKPSYPTTKYKQKPSYPTTKYKSKSYPTTKYKSKKTYPP 350
QY 352 EPSPTTKPEA-PTTPKEPAPTTPKPKP-APT-TPK-----EPAPTTTKEPAPTTTKPK 401
Db 351 TYKPKITYPTTKPKPSYPPSYKPKKITYSTYKPKITYPTTKPKPSYPPSYKPKTYPP 410
QY 402 APTAPKEPAPTTPKETAP-----TTPKKTLPPTTPEKAPTTTPEKAPTTTPEELAPTTPEE 456
Db 411 T-YKPKISYPTTKPKASYVSSYKSKKTYPTTKYKISYPTTKPKPSYPTTKPKTYPP 469
QY 457 PT-PTTPEEPAPTTPKAAAPTPK-EPAPTTTPEKEPAPTTP-----EPAPTT--PKETAP 507
Db 547 PT-PTTPEEPAPTTPKAAAPTPK-EPAPTTTPEKEPAPTTP-----EPAPTT--PKETAP 507

RESULT 11
FPI_MYTGO STANDARD; PRT; 872 AA.
AC Q25434;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR (FOOT PROTEIN 1) (MCFF1).
GN FPI.
OS Mytilus coruscus (Sea mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=42192;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Foot;
RX MEDLINE=96394686; PubMed=8798340;
RA Inoue K., Takeuchi Y., Takeyama S., Yamah E., Yamazaki F., Odo S.,
RA Harayama S.
RT "Adhesive protein cDNA sequence of the mussel Mytilus coruscus and
its evolutionary implications.";

QY 92 TTNKQSTDSGEKTTSAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEP 151
 Db 3 TTAATAAATGKTATAAGSAAABKTA--AGEVSAPPT--AAVAATGEDATT----- 51
 QY 152 TPTTPKEPAPTTPKEPTPTTIKSAPTTPKBPAPTTPKAPTTPKAPTTPKAPTTPK 211
 Db 52 -----AAATAAAEITAAGEAPTTPAPTAAAGKAPTTPAAAGKAPTTPAAAGAPT 103
 QY 212 EPAPTTPKEPAPTTPKTS-APTTPKEPAPTTPKBPAPTTPKAPTTPKAPTTPKAPT 269
 Db 104 GKAPATAAAPVPTTAASKAPTTPAAATHTAAATAAPTTPAAASKKRSSTSSSEEBEHC 163
 QY 270 --PTTPKEPAPT--TPKEPAPTAPKAPTTPK-----EPAPTTPKEPAPTTPKE 314
 Db 164 VKPSKREMGSGITKKQ-----CKKNCCFPKGGHGHCFHFKPGKSHSEHTTTTK- 218
 QY 315 PSPTTPKEPAPTTPKSAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 374
 Db 219 -----APTIIQIATTTT--TPTTT-----TTTKATPTT----- 245
 QY 375 KKPAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPT 434
 Db 246 -----TTTKATPTT-----TTTKATTTT-----TPTT----- 270
 QY 435 PTPKEPAPTTPPELAPTTPPEPTTPPEPAPTTPKAAAPTTPKAPTTPKAPTTP 494
 Db 271 -----TTTTKATPTTTTITP----- 289
 QY 495 KEAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPT 554
 Db 290 -----TTTTKATTTTTSGECKMBPSK-----REDCGYSGITESOCR 328
 QY 555 PKG-----TAPT-----PKAPTTPKAPTTPKGTAPTTLKAPTTPKAPKAPKELA 604
 Db 329 TKGCCFDSISIPQKWCFTLSOVADCKVEPSQVDCGRGIT-----ADQCKNCCFDS 384
 QY 605 PTTTKGPTSTSDKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPT 664
 Db 385 ISGTWKCFYSTQVAA--TKTTTTPTTTTTPTTTTTKATTTT-----TTTTTPTTT 436
 QY 665 KEPTTIHKSFDSTSELSAEPKPALENSPKPGVPTTKPAATKPEMTTAKDKTTERD 724
 Db 437 TTTTT-----TTKA-----TTTTPTTTTTTTTT--KAT--- 463
 QY 725 LRTTPTTTTAAPKMKETATTTTEKTESKITAATTTQVSTTTTQDTPKTTTLKTTTAP 784
 Db 464 -TTTTTTTTT-----TTTTTKAT-----TTTTPTTTTTTTTTKATTTTPTTTT 510
 QY 785 KVTTTKKTTTTTEIMNKPEETAKPKDRATNSKATTPKQKPKAPKPKISTKKPK---T 840
 Db 511 TTTTTKATTTTTSGECKMB-----PSKRADCGYPGITESOCRSGCCFDSISIPQKWC 566
 QY 841 MPRVRKPKTTPR 854
 Db 567 LPQVADCKVAPSSR 580
 RESULT 13
 ID RPB1_HUMAN STANDARD; PRT; 1970 AA.
 AC P24928;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPB1).
 GN POLR2A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92178992; PubMed=1542581;

RA Wintzerith M., Aker J., Vicaire S., Vigneron M., Keding C.;
 RT "Complete sequence of the human RNA polymerase II largest subunit";
 RL Nucleic Acids Res. 20:910-910(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95347616; PubMed=7622068;
 RA Mita K., Tsuji H., Morimyo M., Takahashi E., Neno M.,
 RA Ichimura S., Yamauchi M., Hongo E., Hayashi A.;
 RT "The human gene encoding the largest subunit of RNA polymerase II";
 RL Gene 159:285-286(1995).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
 RNA(N).
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
 CC -1- THE PHOSPHORYLATION ACTIVATES POL2.
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
 FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
 PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
 III FOR 5S AND TRNA GENES.
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC EMBL; X63564; CAA5125.1; .
 DR EMBL; X74874; CAA52862.1; .
 DR EMBL; X74873; CAA52862.1; JOINED.
 DR EMBL; X74872; CAA52862.1; JOINED.
 DR EMBL; X74871; CAA52862.1; JOINED.
 DR EMBL; X74870; CAA52862.1; JOINED.
 DR PIR; S21054; S21054.
 DR MIM; 180660; .
 DR InterPro; IPR000684; RNA_polII_repeat.
 DR InterPro; IPR000722; RNA_pol_A.
 DR InterPro; IPR002879; RNA_pol_A2.
 DR Pfam; PF00623; RNA_pol_A; 1.
 DR Pfam; PF01854; RNA_pol_A2; 1.
 DR PROSITE; PS00115; RNA_POL_II_REPEAT; 43.
 DR Transferrase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
 KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
 FT ZN_FING 71 87
 FT DOMAIN 1590 1958 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
 FT CONFLICT 1067 1067 W -> L (IN REF. 2).
 FT CONFLICT 1449 1449 D -> Y (IN REF. 2).
 FT CONFLICT 1449 1449 D -> Y (IN REF. 2).
 SQ SEQUENCE 1970 AA; 217205 MW; 6876FC25692A657E CRC64;
 Query Match 9.8%; Score 503.5; DB 1; Length 1970;
 Best Local Similarity 33.3%; Pred No. 2e-13;
 Matches 181; Conservative 101; Mismatches 167; Indels 95; Gaps 47;
 QY 151 PTPPTTPKEPAPTTPKE-PTPTTIKSAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKE 204
 Db 1507 PSPMGISPAITPMWQATPAYGAMSPVSGWTPGAGFSPSAASDASGSPGSPAWS 1566
 QY 205 PAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 262
 Db 1567 PTPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSP 1624
 QY 263 TTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPT 322
 Db 1625 TSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSP 1671
 QY 323 PAPTTPKSAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 382

CC DR EMBL; M37227; AAA41693.1; ALT_FRAME.
DR EMBL; X13804; CAA32038.1; ALT_FRAME.
DR EMBL; M21964; AAA41695.1;
DR EMBL; J04517; AAA41692.1;
DR PIR; A30796; A30796.
DR PIR; A25649; A25649.
DR PIR; B25649; B25649.
DR PIR; S02003; S02003.
DR InterPro; IPR001664; IP.
DR Pfam; PF00038; Pfam; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.
FT DOMAIN 1 276 641 51 X 3 AA TANDEM REPEATS OF K-S-P.
FT CONFLICT 164 164 L -> I (IN REF. 2).
FT CONFLICT 185 185 I -> S (IN REF. 2).
FT CONFLICT 193 193 L -> T (IN REF. 2).
FT CONFLICT 199 199 M -> T (IN REF. 2).
FT CONFLICT 346 346 K -> N (IN REF. 1).
FT CONFLICT 373 373 A -> V (IN REF. 1 AND 4).
FT CONFLICT 482 482 G -> E (IN REF. 2 AND 4).
FT CONFLICT 485 485 P -> S (IN REF. 2).
FT CONFLICT 570 571 RK -> KE (IN REF. 2 AND 4).
FT CONFLICT 591 591 P -> T (IN REF. 2 AND 4).
FT CONFLICT 727 727 A -> V (IN REF. 4).
FT CONFLICT 757 759 AAP -> GST (IN REF. 4).
FT CONFLICT 769 769 T -> L (IN REF. 2).
FT CONFLICT 775 775 R -> P (IN REF. 2 AND 4).
SQ SEQUENCE 831 AA; 89486 MW; 1B0973C3F13EF768 CRC64;

Query Match 9.6%; Score 497; DB 1; Length 831;
Best Local Similarity 27.6%; Pred. No. 1.8e-13;
Matches 195; Conservative 59; Mismatches 307; Indels 146; Gaps 35;

QY 37 POTSITQHNKVTSPKLTAKIPNPSPNSDTSKETSIVNKTETVETKETTINK 96
DB 197 PSMST--HIKVKSEKIKVVE-----KSEKEIVVEEOTEETIQVTEED 242
QY 97 TSDGKEKTSKETSIAKDLAPTSKVLAKPTKPAETTTGKPAITTKPEP---T 152
DB 243 KEAQGESEAEGBGEAAATTS-----PPAEEAASP-----EKETSPVKEAKSPAEAKS 293
QY 153 PTPPKPEPA-STTPKEPTPTTIKSAPTTPKEPA---PTTTKSAPTTPKEPA---PTTTK 203
DB 294 PAEAKSPAEAKSPAEVKSAPAEVKSAPAEVKSAPAEAKS-PAEAKSPAEVKSAPTVK 352
QY 204 EP-----APTTPKEPA---PTTTKEPA---PTTTKSAPTTPKEPA---PTTPKKA 245
DB 353 SPGEAKSPAEAKSPAEVKSAPAEAKSPAEAKSPASVKS-PGEAKSPAEAKSPAEVKSAPTV 411
QY 246 --PTTPKEPA---PTTPKEP-----TPTTPKEPAPTTPKEP---APTTPKEPA---PTA 287
DB 412 KSPVEAKSPAEVKSAPTVKSPAEAKSPVEVKSAP-SVKSPSEAKSPAGAKSPAEAKSPVV 470
QY 288 PKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 338
DB 471 AKSPAE--KSPGAKPPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEVKSAPAEAKSP 528
QY 339 APTTTKS-----APTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 389
DB 529 VKGAKSLAEAKSPAEVKSAPVKEIKPPEAKSPAEVKSAPAEAKSPAEAKSPAEAKSPAE 588
QY 390 PKPAPTTPKPAETAPKEPAPTTPKPAETAPKLPKLTPTTPKEPAPTTPKEPAPTTPKEP 449
DB 589 AKPPAEAKSPADIRSPQVKSAPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSP 644
QY 450 APTTPEPTPTTPKEPAPTTPKAAAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 509
DB 645 ---KAKEPPKVEEKTPTAPKTEVKESKDEAPKQKP-KAEKEKPLETKPKDSGEA 700
QY 510 PKGTAPTTLKEPAPTTPKPKKPAKELAPTTPKEPTSTTSKDPAPTTPKGTAPTTPKEPAPT 569

RESULT 15
NFH_RAT
ID NFH_RAT STANDARD; PRT; 831 AA.
AC P16884; Q63368;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).
GN NFH OR NFH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=89065087; PubMed=3143606;
RA Breen K.C., Robinson P.A., Wion D., Anderton B.H.;
RT "partial sequence of the rat heavy neurofilament polypeptide (NF-H).
RT Identification of putative phosphorylation sites.";
RL FEBS Lett. 241:213-218(1988).
RN [2]
RP SEQUENCE OF 37-831 FROM N.A.
RX MEDLINE=88309090; PubMed=2457365;
RA Dautigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,
RA Jolles P.;
RT "the large neurofilament subunit (NF-H) of the rat: cDNA cloning and
RT in situ detection.";
RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
RN [3]
RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.
RX MEDLINE=87080760; PubMed=2878828;
RA Robinson P.A., Wion D., Anderton B.H.;
RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide
RT (NF-H).";
RL FEBS Lett. 209:203-205(1986).
RN [4]
RP SEQUENCE OF 318-831 FROM N.A.
RX MEDLINE=89184647; PubMed=2928342;
RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D.,
RA Smulowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;
RT "Cloning of a cDNA encoding the rat high molecular weight
RT neurofilament peptide (NF-H): developmental and tissue expression in
RT the rat, and mapping of its human homologue to chromosomes 1 and
RT 22.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC

```
Db 701 KKEEA-----KEKKAAPPEETPAKLGVKEEAKPKEKAEDAKA-----KEPSKP 744
Qy 570 TPKEPAPTTPKGTAPTTILKEPAPTTPKKPAKPAKELAPTTTKGPTSTTSDDKPAPTTPKETAP 629
Db 745 SEKE-----KPK-----KEEVPAAPEKKDTKE-----ERTTESKKREKPKMEAK 784
Qy 630 TTPKEPAPTTPKKPAPTTPPETPPPTTSEVSTPTTTTKEPTTIHKSPDE 676
Db 785 AKED-----KGLPQEPSKPKTEKAEKSSSTDQKDSQPSEKAPED 824
```

Search completed: April 26, 2002, 16:19:35
Job time: 360 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:16:52 ; Search time 49.78 Seconds
(without alignments)
437.589 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1167
Perfect score: 5155
Sequence: 1 VKDNKNRTRKKRTPKPPV.....GKPDGLTLRLNCTLVAFRG 968

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	522	10.1	805	4	US-09-103-429A-4
2	506.5	9.8	744	6	5202236-25
3	504	9.8	1837	3	US-08-928-361B-5
4	498.5	9.7	786	4	US-09-103-429A-3
5	489	9.5	1721	3	US-08-700-651-5
6	489	9.5	1721	3	US-08-928-361B-6
7	488.5	9.5	826	1	US-07-638-431-2
8	488.5	9.5	826	5	PCT-US92-00018-2
9	476.5	9.2	652	6	5202236-13
10	452	8.8	960	4	US-09-219-849-5
11	423.5	8.2	1867	2	US-08-479-537A-5
12	423.5	8.2	1867	4	US-09-083-116-5
13	423.5	8.2	2035	2	US-08-479-537A-2
14	423.5	8.2	2035	4	US-09-083-116-2
15	419.5	8.1	2476	2	US-08-276-967-2
16	419	8.1	1185	4	US-09-041-886-23
17	417	8.1	829	1	US-08-642-255-132
18	417	8.1	829	1	US-08-397-633A-53
19	417	8.1	837	1	US-08-175-155-68
20	417	8.1	837	1	US-08-477-509B-103
21	417	8.1	837	1	US-08-642-255-101
22	417	8.1	837	2	US-08-707-237A-75
23	417	8.1	837	3	US-08-482-085B-103
24	417	8.1	897	1	US-08-397-633A-50
25	413.5	8.0	907	3	US-08-783-774-2
26	413.5	8.0	907	5	PCT-US95-04611A-19
27	404.5	7.8	408	1	US-07-609-716-65

28	404.5	7.8	408	4	US-08-475-411A-65	Sequence 65, Appl
29	404.5	7.8	408	4	US-08-478-029A-65	Sequence 65, Appl
30	398.5	7.7	682	1	US-08-642-255-126	Sequence 126, App
31	398.5	7.7	682	1	US-08-397-633A-36	Sequence 36, Appl
32	381	7.4	1848	4	US-08-296-791-6	Sequence 6, Appl
33	381	7.4	1848	5	PCT-US95-10661A-6	Sequence 6, Appl
34	376	7.3	960	4	US-09-219-849-6	Sequence 2, Appl
35	374.5	7.2	1537	1	US-08-325-267A-2	Sequence 114, Appl
36	371.5	7.2	1231	3	US-08-904-263A-4	Sequence 114, Appl
37	371	7.2	761	2	US-08-707-237A-84	Sequence 4, Appl
38	371	7.2	762	1	US-08-642-255-114	Sequence 84, Appl
39	371	7.2	762	1	US-08-397-633A-26	Sequence 114, Appl
40	370.5	7.2	1084	1	US-08-642-255-62	Sequence 62, Appl
41	368	7.1	762	1	US-08-642-255-120	Sequence 120, App
42	368	7.1	762	1	US-08-397-633A-31	Sequence 31, Appl
43	367.5	7.1	1187	1	US-08-320-559-28	Sequence 28, Appl
44	367.5	7.1	1187	3	US-08-545-860D-28	Sequence 28, Appl
45	367.5	7.1	1187	5	PCT-US94-04496-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-09-103-429A-4
; Sequence 4, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang, Ping
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
; TITLE OF INVENTION: CDNA and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
; STREET: 118 No. 6187558th Tioga
; CITY: Ithaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103.429A
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-2000
; TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 805 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; TISSUE TYPE: peritrophic membrane
US-09-103-429A-4

Query Match 10.1%; Score 522; DB 4; Length 805;
Best Local Similarity 27.8%; Pred. No. 5.2e-26;

FILING DATE: 24-NOV-1986
APPLICATION NUMBER: 650,128
FILING DATE: 13-SEP-1984
SEQ ID NO: 25
LENGTH: 744
5202236-25

Matches 226; Conservative 33; Mismatches 211; Indels 342; Gaps 40;

QY	125	TSKVLAKPTPKAETTTGKALITPKETPTTTPKEPASTTPKEPTPTTTPKTSAP-----TTP	179
Db	9	TALGLVAARPEVSDAEKFNALHPHPDXP--FAEQXLLLPXEYDCTKFYYCEYGLKFAP	66
QY	180	KEPAPTT-----TKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK	227
Db	67	RDCAPTGTEFKESAQTCVHAALAGCTLPGPAET---QAPATQ---APTITQ---APTITTT	119
QY	228	SAPTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTA	287
Db	120	QAPTTTQAATTT-----QAPTTQAATTT-----QATTTQAATTT-----	156
QY	288	PKRAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAP	347
Db	157	---QAPTTTQ---APTITQ---APTITQ---APTITQ---APTITQ---APTITQ---APTITQ	198
QY	348	TTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK	391
Db	199	TTPAATTPAAT--TPAATTPAATTPAATTPGVPAPTS-----APVWPICELLPGNCADFD	253
QY	392	-----EPAPTTTPK	399
Db	254	IHLIPHDKYCNLFYQCSNGYTFEQRCEGLYFNPIYVQRCDSANVECDGEISPAFPVIE	313
QY	400	K-----PA	402
Db	314	GNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDDKYQCVHGNLVERRCGAGTHFSFELQ	373
QY	403	-----PTAP--KEPAPT--	417
Db	374	QCDHIELWGCTLPGGESEVDVDEDACDAGWYCPTEPIEWEPPLNGCPADFSIDHLLPHES	433
QY	418	-----APTTPKKLTPTT-----PEKLAFTTPKEPAP	443
Db	434	DCGQYLCQVHGQTIARPCGNLHSPATQSCSPVTAGCQVFECDSNQCTSTAAPTAAP	493
QY	444	TTPEELAPTTPEPTTPKEPAPTTTPKAAAPNTPKAPNTPKAPNTPKAPNTPKAPNTPK	503
Db	494	TAAPTAAPTAAPTAAPSPVVPVPA--TPATAAPVVPVPTAAPT-----PAPTAAPTAAPTA	548
QY	504	ETAPTT---PKGAPTTLKEPAPTTTPKPKAPKELAPTTTKEPTSTSDKAPTTTPKGATP	560
Db	549	PSPTTVVVPVPAFTA---APT---AVPEIPTVTSAPTAAPT---AAPTAAPTAAAP	598
QY	561	TPKEPAPTTTPKEPAPTTTPKGTATTLKEPAPTTTPKPKAPKELAPTTTKEPTSTSDKAP	619
Db	599	TTAVPEIPTT-----VTSPTTAAPT--AAPANT-----TVPVPTAAPTAAAP	641
QY	620	APTTPKEAPTTTPKEPAPTTTPKPKAPTTTPPTTPPTTSEVSTPTTKEPTTLHKSPDESTP	679
Db	642	AP-----NTTVTAAPTAAPTT--AAPANTTVVVPVPTAAPTAAAP-----PTVAH-----AP	685
QY	680	ELSAEPTPKALENSKPEGVPTTKTPAATKPE	711
Db	686	NITAAP-----VTTTSAPANT-PE	703

Query Match 9.8%; Score 506.5; DB 6; Length 744;
Best Local Similarity 29.9%; Pred. No. 4.6e-25;
Matches 239; Conservative 78; Mismatches 308; Indels 175; Gaps 50;
QY 51 PKTTTAKPINRPSUPPNSDTSKETSITVNKETTETVETKETTITNKQSTDCGKETTISAKE 110
Db 24 PKMTYPPYTKPSPYPP-----IVKSKPTY-----KPKIT----- 53
QY 111 TOSIEKTSAKDLAPTSKVLAKPTKAEITTTKGPALTTPKEPTPTTPKE-----PASTTPKE 166
Db 54 -----YPPTYK--AKPS-----YPPTYKPKKTYPPPYTKPKLTYPPTYKPKP 92
QY 167 PTPTTIKSAPT--TPKEPAPTTTKSAPTTTPKBPAPTTTPKEPAPTTTPKEPAPTTTPKEA--P 223
Db 93 SYPPTYKSKPTTKPKLITPPYKAKPSYPPTYKPKAKTYPPPT--YKPKLTYPPTYKPKASYP 151
QY 224 TTTKSAPTTTPKEPAPTTTPKAPPT--TPKEPAPTT-----PKETPTTPKEPAPTTKEP--- 275
Db 152 PTKPKPSYP--PSYTKTKTYPPTYKPKLTYPPTYKPKPSYPPSYKPKKTYPPTYKPKLT 209
QY 276 -APTTPKEA--PTAPKPA--PTTPKEA--PTTPKEA--PTTPKEA--PTTPKEA--PTTPKEA 324
Db 210 YPPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPS 269
QY 325 -PTTKSAPTTTPKEPAPTTTKSAPTTTPKPSPTTTPKEA--PTTPKEA--PTTPKEA--PTTPKEA 377
Db 270 YSTYKAKPSYPT--YPTYKAKPSYPT--PTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 322
QY 378 A--PTTPKEA--PTTPKEA--PTTPKEA--PTAPKEPAPTTTPKEPAPTTTPKLTPTT 429
Db 323 SYPPTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPS 382
QY 430 PEKLAFTTP--KPAPTTPPELAPTTPEPT--PTTPKEA--PTTPKEA--PTTPKEA--PTTPKEA 478
Db 383 TYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 438
QY 479 KEAPTTTPKEPAPTT-----PKAPTTTPKEPAPTT--PKGTAPTTTPKEPAPTTTPKPKAPK 532
Db 439 SYPPTYKPKLISYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 498
QY 533 ELAPTTTKEPTSTSDKAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTTPKEA-- 591
Db 499 SYPPTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPS 546
QY 592 -PTTPKPKAPKELAPTTTKGPT--STTSOKAPDTTPKEPAPTTTPKEPAPTTTPKPA--PT 646
Db 547 PPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 601
QY 647 ---TPETPTTPPTTSEVSTPTTKE-----PTTIHKSPD-----ESTPELSAAPT--PKALENS 693
Db 602 YKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAK 661
QY 694 KPEGVPTTKTPAA-----TKPEMTTTAKDKTTERDLRTPPTTTAAAP-----KMT 739
Db 662 PSYP--PTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAK 719
QY 740 KETATTTTEKTESKITATTT 759
Db 720 YKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 719

RESULT 3
US-08-928-361B-5
; Patent No. 6071518
; Sequence 5, Application US/08928361B
; Patent No. 6071518
; Title of Invention: METHOD OF PRODUCING BIOADHESIVE
; Protein
; Number of Sequences: 39
; Current Application Data:
; Application Number: US/07/528,762
; Filing Date: 25-MAY-1990
; Application Number: 82,456
; Filing Date: 07-AUG-1987
; Application Number: 933,945

GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928.361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026.062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: VERNY, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1837 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-5

Query Match 9.8% Score 504; DB 3; Length 1837;
Best Local Similarity 22.3%; Pred. No. 1.8e-24;
Matches 279; Conservative 81; Mismatches 481; Indels 412; Gaps 40;

QY 55 TAKDINPRPSLPPNSDTSKTSLVNKET-----TVETKETTTNKQTST-----DGK- 102
DB 215 TGLPTDYPNCPNPVTGNLVSRSRGTKIPNTYAGVYRSNETKTTSPSANTNELLVDPKI 274
QY 103 -----EKTTSAKETQSIKTSKDLAPTISKVLAKPTPKAETTTKGPALTTTPKEPTTPPK 158
DB 275 NAPCNSENSEQVOIFDMGSKVYIPYTKCVGVKHTTTTTT-----TTTTTTTTTTTTT 329
QY 159 PASTTPKEPTPTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 218
DB 330 TTTTNTTTTTTTTTTTTTTTT-----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 383
QY 219 KEPAPTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 278
DB 384 TTTTNTTTTTTTTTTTTTTTT-----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 443
QY 279 TPKEPAPTAPKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 338
DB 444 TTTTNTTTTTTKKPTTT-----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 497
QY 339 APPTTSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 376
DB 498 TTTTNTTTTTTKKPTTTTATTTTSETSVIKPDENCWLEKNCCEAKGATVGVIG 557
QY 377 ----- 376

Db 558 KGRINGMAFTMIPNDTHVRFKVKVDGNTISVRCGKGAGKLEFPDRSLDFTIPPVA 617
QY 377 -----PAPTPE-----PAPTPE-----PAPTPE-----PAPTPE-----PAPTPE 388
Db 618 GHNSCSIIIVSGGGKTHVSPYSGKDVLSAPLOCELEFNEVYCDTCTAKYGAHSGYQ 677
QY 389 TPKEPAPTTPKPAPTAPKAPETTPKETAPTTPKLTTPTPPEKLPAPTTPPEKLPAPTTPPE 448
Db 678 TSADFVTTTAKPTTTT--TCAPCQPTTTTGTGSPKPTTTTTPKATTTT-----TILNP 729
QY 449 LAPTPPEPTTPPEPAPTTPKAAAPNTPKAPPTPKAPPTTPKEPAPTTPKEPAPTTPKEPAPT 508
Db 730 IITTTOKPTTTT-----TT-----KVPGRPIATTITTLKPIVITTTTTPKATTTTTPVT 780
QY 509 TPKGAPTTLKEPAPTTPKPKAPK-----ELAPTTTKEPTS-----TTSOK 549
Db 781 T-----TTTTKRDEMTTTTPLDIDGIEITPIPIEKMLDKYTRMIYDYNGLLLDSNDE 835
QY 550 PAPTTPKG-TAPTTPKEPA-----PTTPKEPAPTTPK----- 580
Db 836 PIPGSAQGIADTSNLFPPVQTHKSTGLPIDPMVGLPDPKSGNLVHPYTNQTMGSLSVY 895
QY 581 -----GTAPTTLK-----EPAPTTPKPKAPKELAPTTT----- 608
Db 896 LAAKNLTVDDETYGLPIDTLGYPLDPVSLIPFPNETGELFDPISDEIMNGTIAGIVSG 955
QY 609 -KGPTSTTSKPAPTTPKE-----TAPTTPKEPAPTTPKPAPTTPETP----- 651
Db 956 ISASELSLQKSAPIDPATNMVVGEGGLNPNATGMVPGSLGPSQTPFSEIEDGGII 1015
QY 652 PPTTSEVSTPTTKEPTTIHKSPDESTPE-----LSAEPK 688
Db 1016 PP---EVAANADKFKLSIPSPVPSIPEKQKIDISSELMYDIESGRUIGOVSKRPIPG 1072
QY 689 ALENSKPEGVPTTKTPAATKPEMTTAKDKTTERDLRTPTTETTAAPKMTKETA---T 744
Db 1073 SIAGDLN-----PIMKTPTQT-----DSVTGKPI-----DPTTGLP-FNPPTGHLNP 1114
QY 745 TTEKTTESKITATTTQVSTTTQDTPPTPKITTLTKTTTTLAPK---VTTTKKTTITTEIMNKP 802
Db 1115 TNNNTMDSFAGAYKAVSNGIKTDNVYGLPVEITGL-PKDPVSDIPFNSNTGELVD-- 1171
QY 803 EETAKPKDRAT-----NSKATTPKQKPTKAPKPTSTKPKTMPVRK 846
Db 1172 PSTGKPINNYTAGIVSGKRLGPPIDENGNLFPDPTKLPIDGNOLVAPENSTSVSGSTS 1231
QY 847 PKTTTPPKMTS---TMPE-----LNPTSRIAEAMLOTTTTPRNPQTPSKL 888
Db 1232 GSTKPKGIPVNGGVVPEDEAKDQADKDGKDLIVPPTNSINKDPVTNTQYSNTTGNI-- 1289
QY 889 VEVNPKSEDAGGAGCETPHMLLRPHVFMPEVTPD-----MDYLPVPVNGQIINP-- 938
Db 1290 --INF--ETGKVPISGLPSGLNYSFNTPOQTDEITGKVPDVTVTGLPVPDSTGEIIDPAT 1345
QY 939 -----MUSDETNICNGK-----PVDGLTTLRNGTLV 964
Db 1346 KLPIPGSVAGDEILTEVLNITTDVETGLPIDLETGLPRDPVSGLPQPNGLTV 1398

RESULT 4
US-09-103-429A-3
; Sequence 3, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang, Ping
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
; TITLE OF INVENTION: CDNA and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
; STREET: 118 No. 6187558th Tloga
; CITY: Ithaca


```
Db 467 VKDVGNTISVRCRKGAGKLEFPDRSLDFTTPVAGHNSCSIIIVSGDGKIHVSGSKD 526
QY 369 PAPTTPKKPAPTTPKE-----PAPTTPKEPAPTTPKKPAPTAPKPA 410
Db 527 VSLIS-----APIQSELFNEVYCDTCTAKYGAHSGYQTSADRVTTTAKPTTTT--TGA 580
QY 411 PTPKETAPTTPKLTPTTPEKLAFTPEKPAFTPEELAPTPEPTTPEEPAPTTP 470
Db 581 PCQPTTTTSGSPSKPTTTTAKATTT-----TTLNPIITTTTQKPTTT-----TT- 627
QY 471 KAAPNTPEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKPA 530
Db 628 --KVPCKPIATTTTTLKBIIVTTTAKATTTTTPVT-----TTTKROEMTTTTL 680
QY 531 PK-----ELAPTTPKEPT-----TTSDKPAFTTPKG-TAPTTPKEPAPT 570
Db 681 PDIGDIEITPIERKMLDKYTRMIDYNSGLLSDNDEPIFGSQAGQIADTSNLFVQTH 740
QY 571 PKEPAPTTPKGAPTTLKE---PAPTTPKPAKELAPTTPKGTSTTSKPAFTTPKET 627
Db 741 KSTGLPIDPMVGLPDPKSGNLVHPYTNQWMSGLSVLAAKNLVTDITYG--LPIDT 798
QY 628 APTTPKEPAPTTPKPAFTTPPTTSEVSTPT-----TTKPTTHKS----- 673
Db 799 LTGYPLDPVSLIPFN--PETGELEFDPISEIMNGTIAGIVSGISASESLLSOKSALIDPA 856
QY 674 -----PDEST-----KQKIDISSELMYDIESGRIGOVSKRPIPGSIAGDLNP 963
Db 857 TNNVYGERGLLNPATGVMIPGFLGPSQOTQFSEIEDGGIIPPEVAAANADFKLSIP- 915
QY 696 EPGVPTTKTPAATKPEMTTAKDKTTER-----DLRTPPTTTAAPKM 738
Db 916 -PSVP-----ESIP-----KQKIDISSELMYDIESGRIGOVSKRPIPGSIAGDLNP 963
QY 739 TKETATITTEKTESKITATT--TQVTSITTQDTPFKITLTKTTTLAPKVTTTKTITTE 797
Db 964 IMKPTQDTSVTGPIDPTTGLPNPTGHLINPTNNMTDSSFAGAYKAVANGIKTDN 1023
QY 798 IMNKP--EETAKPD-----RAINSKATTPKPKPTKAPKPTSTTKPKTMP----- 842
Db 1024 VIGLVPGEITGLPKDPSDIPFNSTTGELVDPTGKPIINNSTAGIVSGRGLPIEDENG 1083
QY 843 -----RYRKPKTTPTPKMTS--TMPE----- 862
Db 1084 NLEDPSTNLPIDGNQLVNETSTVSGSTGTTKPKPIPVNGGVVPEEAKDQADKG 1143
QY 863 -----LNPTSRIAEMLOTTTRNPQTPNSKLVENPKSDAGGAEGETPHMLLRPHVMP 917
Db 1144 KGLIVPPTNSINKDPVTNTQYSNTGNI-----INP--ETGKVIPLGSLGSLNYPSFNTP 1197
QY 918 EVTPD-----MDVLPVPNOGIILNP-----MLSDETNICNGK----- 950
Db 1198 QQTDEITGKPDVITVGLPVDPTGGEIIDPATKLPISGVAGDEILTEVLNITDVTGLP 1257
QY 951 -----PVDGLTTLRNGTILV 964
Db 1258 IDLETGLPDPVPSGLPQLPENGTLV 1281
```

RESULT 6

US-08-928-361B-6

; Sequence 6, Application US/08928361B

; Patent No. 6071518

; GENERAL INFORMATION:

; APPLICANT: Petersen, Carolyn

; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,

; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS

; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM

; TITLE OF INVENTION: SPECIES INFECTIONS

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PETERS, VERNY, JONES & BIKSA

```
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verny, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1721 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-6
```

Query Match

9.5%; Score 489; DB 3; Length 1721;

Best Local Similarity 22.1%; Pred. No. 1.5e-23;

Matches 284; Conservative 82; Mismatches 476; Indels 442; Gaps 41;

```
QY 5 KNRTRKPKPTPKPPVVDGASGLDNGFKVTPDTSTTQHNVKSTSPKITTAKPINRPS 64
Db 116 RSNETH---TTPEPSANTYAGVYRSN-ETKTPEPSANT---NFLVDPKI-----N 158
QY 65 LPPNSDTSKE-----TSLVNETTVETKETTNTNKTSTDGKEKTTSAK 109
Db 159 APCSENSEFGQGFDMGSKVYIPYTKCVGVKHTTTTTTTTTTTTTTTTTTTT----- 208
QY 110 EQSIEKTSADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPASTTPKEPTP 169
Db 209 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 240
QY 170 TTIKSAPTTKEPAPTTPKSAPTTPKEPAPTTPKAPTTPKEPAPTTPKAPTTPKSA 229
Db 241 TTTTTPPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 300
QY 230 PTTTPKEPAPTTPKPKAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTAPK 289
Db 301 TTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 350
QY 290 KPAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPT 349
Db 351 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 406
QY 350 PKB-----PSPTT----- 357
Db 407 TSETSVIKPDEWCWLEKNGECEAKGATVYGVIGKDRIENGMAFTMIPNDTHVRFREK 466
QY 358 -----TKBP-----APTTPKE 368
Db 467 VKDVGNTISVRCRKGAGKLEFPDRSLDFTTPVAGHNSCSIIIVSGDGKIHVSGSKD 526
QY 369 PAPTTPKKPAPTTPKE-----PAPTTPKEPAPTTPKPAFTAPKPA 410
Db 527 VSLIS-----APIQSELFNEVYCDTCTAKYGAHSGYQTSADRVTTTAKPTTTT--TGA 580
```

RESULT 8
PCT-US92-00018-2
: Sequence 2, Application PC/TUS9200018

7
 RESULT US-07-638-431-2
 ; Sequence 2, Application US/07638431
 ; Patent No. 5198535
 ; GENERAL INFORMATION:
 ; APPLICANT: Hoffman, Stephen L.
 ; APPLICANT: Charoenvit, Yupin
 ; APPLICANT: Hedstrom, Richard
 ; APPLICANT: Khushmith, Sisin
 ; APPLICANT: Rogers IV, William O.
 ; TITLE OF INVENTION: Protective malaria sporozoite surface protein
 ; TITLE OF INVENTION: immunogen and gene
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: A. David Spevack
 ; STREET: NMDC Building 1 T-12 National Naval
 ; STREET: Medical Center
 ; CITY: Bethesda
 ; STATE: MD

```

GENERAL INFORMATION:
APPLICANT: Hoffman, Stephen L.
APPLICANT: Charoenvilt, Yupin
APPLICANT: Hedstrom, Richard
APPLICANT: Khumwath, Srisin
APPLICANT: Rogers IV, William O.
TITLE OF INVENTION: Protective malaria sporozoite surface protein
TITLE OF INVENTION: Immunogen and gene encoding
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESS: A. David Spevack
STREET: NMRC Building 1 T-12 National Naval
STREET: Medical Center
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20814-5044
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00018
FILING DATE: 19920103
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, Avram D.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6755
TELEFAX: (301) 295-4033
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 826 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-00018-2

```

```

Query Match          9.5%  Score 488.5;  DB 5;  Length 826;
Best Local Similarity 26.4%  Pred. No. 7.2e-24;
Matches 153;  Conservative 54;  Mismatches 223;  Indels 149;  Gaps 34;

QY 168 TPTTISAPTTKEPA--PTTISAPTTKEPA--PTTKEP---APTTPKEP--APTTPKE 220
DB 273 TCKVNDQCPQIPVPIPKIKEKPSNPEEPVNPNDPNPNNNNNNNNNNNPNPN 332
QY 221 PAPTTSKAPTTKEP--APTTPKEP--APTTPKEP--TPTTPKEPAPTTKEPA 276
DB 333 PNNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNH 392
QY 277 PTTTPKEPAATPAK--KPAATTPKEPAATTPKEPAPTTPKEPAPTTPPTTSAPT 334
DB 393 RNNPKRNNNNKPKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPN 443
QY 335 TTEPATTTKSAPTTPKEPSPTTTPKEPAATTPKEPA--PTTPKPK--APTTPKEAPTTPKE 392
DB 444 PKKPNP-----NPNPNPNKPNPN-----EPNPNPNPNPNPNPNPNPNPNPNPN 492
QY 393 PAPTTPKKAAPPAKPEP--APTTPKEAPTTPKPKLPTTTPKEKLAAPTTPKEKPAAPTTPKE 451
DB 493 PS-----NPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 523
QY 452 TTEPEPTPTTPEEPAPTTPKKAAPTTPKEP--APTTPKEPAATTPKEPAATTPPTTP 510
DB 524 SNPNP-----PLNPNP-----SNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 564
QY 511 KGTAPPTLKEPAPTTPKKAAPKELAPTTPKEPTSTSDKPAAPTTPKGTAPTTPKEAPT 570
DB 565 -----PSNDEEPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 606
QY 571 PKPAAPTTPKGTAPTTPKKAAPTTPKKAAPKELAPTTPKGTAPTTPKEAPTTPKEAPT 630

```

```

DB 607 KPEPI-----NPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 649
QY 631 TPEKP-----APTTPKKAAPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTP 684
DB 650 LPTLPQKGNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 706
QY 685 PTPKALENSPK-----EPGVTTK 703
DB 707 PTKGHEERLPKPRNSDDVYVNNVNNKNNKNDPEEIPNNE 745

RESULT 9
5202236-13
Patent No. 5202236
APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
SUSAN L.; MCCANDLISH, RUSS-WEI, TENA; FILIPULA, DAVID
TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
PROTEIN
NUMBER OF SEQUENCES: 39
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/528,762
FILING DATE: 25-MAY-1990
APPLICATION NUMBER: 82,456
FILING DATE: 07-AUG-1987
APPLICATION NUMBER: 933,945
FILING DATE: 24-NOV-1986
APPLICATION NUMBER: 650,128
FILING DATE: 13-SEP-1984
SEQ ID NO: 13
LENGTH: 652
5202236-13

```

```

Query Match          9.2%  Score 476.5;  DB 6;  Length 652;
Best Local Similarity 29.7%  Pred. No. 3.3e-23;
Matches 221;  Conservative 71;  Mismatches 284;  Indels 169;  Gaps 45;

QY 51 PTTTAKPINRPSLPNNDTSKETSIVNKKETVETTTTNNKOTSDGKEKTSKAKE 110
DB 11 PTKTTPPTTPKPKPSYPP-----TKSKPTV-----KPKIT----- 40
QY 111 TOSIEKTSADLAPSKVLAKEPTPKAETTTKGPALTTPTEPTTPKE-----PASTTPKE 166
DB 41 -----YPTTPK--AKPS-----YPTTPKPKTTPPTTPKKTTPPTTPKPK 79
QY 167 PPTTPKSAPT--TPKEPAATTPKSAPTTPKEPAATTPKEPAATTPKKEPA--P 223
DB 80 SYPTPKSKPTPKITVPTPKAKPSYPTPKPKTTPPT--YKPKLVPPTPKKASYP 138
QY 224 TTTKSAPTTPKPAATTPKKAAPT--TPKEPAATTP-----PKETPTTPKKEPAATTPKEPA 278
DB 139 PTKKPKPSIP--PSYKTKTTPPTPKLTPPTPKPKPSYPSYKPKTYPPTPKPKLT 196
QY 279 TPKEPAATPAKKA--PTTPKKEPA--PTTPKKEPAATTPKKEPSPTTP-----KEPAATTPK 329
DB 197 YP-----PTPKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPSYPSLAKKPSYPTPK 252
QY 330 SAPTTTKEPA-----PTTPKSAPTTPKEPSPTTPKKEPA--PTTPKKEPA--PTTPKPK 376
DB 253 AKPTPKAKPTPKAKPTYSYTKAKPTYP-----PTPKAKPSYPTPKAKPTPKAKPT 308
QY 377 PA--PTTPKKEPA--PTTPKKEPAATTPKKAAPTTPKKEPAATTPKKTTPPTTEK 432
DB 309 PSYPTPKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPSYPT 362
QY 433 LAPTPP-----KPAATTPKEELAPTTPEEPT--PTTPEEPA--PTTPKKAAPTTPKEPAATTPK 487
DB 363 AKPTPKPSYTKKPSYPTPKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPSYPT 413
QY 488 EAP--PTT-----PKPAATTPKKEAPT--PKGTAPTTTKEPAATTPKKAAPKELAPTTPK 540
DB 414 KPTYPSTYKARPSYPTPKPKISYPTPKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPSYPTPK 473

```

OY	541	EPTSTOSKAPVTPKCGAPTAPTEPEAPVTPKCEAPVTIKGVAFTILKEBA--PTTPKKP	598
Db	474	KPTTPTSTYKAKPYTK--AKPSYP-----PTYKAKPS-----YEPTTKAKPSYPTTYKAP	521
OY	599	APKEIAPTTTGKPNSTTSDBKAPATPPKETAAPTPKEBA--PTTPKKPA-PTT---	PETP 651
Db	522	TYK-AKPTYKAKPSYPTTYKAKPSYP---PTYKAKPSYPTTYKAKPTYSTYKAKPSYP	576
OY	652	PPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAETPKALNSKREQVPTTKTPAAKTPE	711
Db	577	PTYKAKPSYPT-----YKAKPSYPTTYKAKPTYKA-----KPYTPT-----YKAKPS	620
OY	712	MUTTAOKDTERDLRTTPETTTAAP	736
Db	621	YPTTHKAKPTYKAKPSYPTTHKAP	645

```

RESULT 10
US-09-219-849-5
Sequence 5, Application US/09219849
Patent No. 6150081
GENERAL INFORMATION:
APPLICANT: VAN HEERDE, GEORGE V.
APPLICANT: VAN RIJN, ALEXIS C.
APPLICANT: BOUWSTRA, JAN B.
APPLICANT: DE WOLF, FREDERIK A.
APPLICANT: MOOBROEK, ANDREAS
APPLICANT: WERTEN, MARC W.T.
APPLICANT: WIND, RICHEL D.
APPLICANT: VAN DEN BOSCH, TANJA J.
TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THERE
TITLE OF INVENTION: PREPARATION THEREOF
FILE REFERENCE: 2728-2
CURRENT APPLICATION NUMBER: US/09/219,849
CURRENT FILING DATE: 1998-12-23
NUMBER OF SEQ ID NOS: 50
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 960
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Illustrative
US-09-219-849-5

```

	Query Match	8.8%	Score 452;	DB 4:	Length 960;	
	Best Local Similarity	25.6%;	Pred. No. 1.8e-21;			
	Matches 173; Conservative	60;	Mismatches 309;	Indels 134;	Gaps 41;	
QY	132 PPKAETTTKGCALPTTKPEPTTTPKEPASTPKEEPTTIKSAPTTKEPAEP--TTTKS	189	:	:	:	:
Db	72 PGAPGPGSGDGG--PPGAPGPAGP--PGSRDDGPRGGAGC-PAGPRGSSNDPPGGAPGA	126	:	:	:	:
QY	190 APTTPKEPAPTITTKEDA--PTTKEPAPTITTKEDAPTITTSAPTTKEPAPTTP--KKPA	245	:	:	:	:
Db	127 GPPGSDPPPGAPGAPGPGSSRDPPPGAPGAPGAGPSRDPGPPGAPBPAGCPGGRDGC	186	:	:	:	:
QY	246 PTTPKEPAPTTP---KEPPTTPKEDA--PTIKEPAPTTPKEDA-----	284	:	:	:	:
Db	187 P--PGAPGAGPGRGSSRDPGPGAPGAGPGRGSDPDPBGHNGPAGKGANGAPRGAGHG	244	:	:	:	:
QY	285 -----PLAK-KPAPTTP---KEPAPTTPKEDA--PTTKEPAPTTPKEPAPTITKS	330	:	:	:	:
Db	245 PAGPKAGHPAGCKGAPGAGGAPGSSDPPPGAPGAGGAPGSSRDGP--PGATGP----A	298	:	:	:	:
QY	331 APPTTKEPAPTITTSAPTTKEPSPPTTTPKEPAPTITTKEDAPTTP---KKPAPTTPKEPA	387	:	:	:	:
Db	299 GPFGSDPPG-----PGAPGAPGAPGSRDPCG--PGATGPAGPGRGSRDGP--PGAFGT	348	:	:	:	:

```

QY 388 TTP---KEPAPTTTKKPA-----PTAPKEPAPTTTEKETAPTTTKULPTTEKLAFTT 437
Db 349 AGPPGSRDPPGAGAPGPPGSGRDPPGAPGAGPPGSGHDCP----PGAGPPAGPPG 404
QY 438 PEKPAPT--TEBELAPTTPEEPPTTPTEBPAPTTPKAA--PNTKEPAPTTP--KEPAPT 432
Db 405 SRDQPPAPAPPPGSRDPCP--PGAPGAGPPGSRDQPPGAPGAPPPGSRDGP 462
QY 493 TPKEPA-----PTTKKET*APTTPKGT-----APTTLKRPATTTTKK*APRELA 535
Db 463 GAHGAGAGKGAHGAGPAGAHGAPGAGHAPGAPKAGPAGPAGPSRDGPAPAPAG 522
QY 536 PTTTKEPTSTSDKPAPTTPKGA--PTTKKEPAPTTP---KEPAPTTPKGA--PTTLKEP 590
Db 533 PPGSGDPPPGCA--PGAPGPPGSDPCPPGAPGAGPPGSGRDPCPAPGAPGPPGSRDP 560
QY 591 APPTTKKPAP-----KELAPTTTKGPTST--TSDKPAPTTPKETAAPTTTEKAPATPK 642
Db 581 GP--PGAGPAPGPPGSRDPPGPPAGPAPGPPGSGNDPPGPPAGPAGPPGSGNDPCP--PGA 636
QY 643 PAPPTP---ETPPPTTSEVSTPTTKKEPTTIHKSPPDSBPBELSAEPTPALEN-----S 693
Db 637 PCPAPPPGSRDPPGAGAPGAGPPGSGRDPCP--PGAPGAGPPGSGRDGPAPAPGAPGPG 695
QY 694 PKEPGVPTTKTPAATK 709
Db 696 SNDGPAPGAGGAPK 711

```

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525


```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note="The amino acids spanning
OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note="Amino acid 134 is X1 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note="Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note="Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1 21
OTHER INFORMATION: /note="Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-08-479-537A-5

```

```

Query Match      8.2%  Score 423.5; DB 2; Length 1867;
Best Local Similarity 24.4%; Pred. No. 2.4e-19;
Matches 235; Conservative 63; Mismatches 461; Indels 203; Gaps 48;

75 TSLVANKETTVETKTTTNTKQSTGDKETTSKAKETOSIEKTSKD-LAPTSKVLAKPT 133
16 TLTV-----VTCGSHASSTPGGEKETSATGRSSVSSTENKNAVSMTSVLSHS 65
134 P-KAETTTKGP--ALTTPKEP-----TPTTPKEPASTTPKEPTPTTIKSAP 176
66 PEGSGSTTGGDVTPLAPATEPASGSAATWGODVTSVPTTRPALGSTPPPAHDVT---SAP 122
177 TTPKEPAPT-----KSAPTTPKEP-----APTTPKEPAPTTPKEPAPT 217
123 --DNKPAFGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 180
218 TKE--PAPTTT-----KSAPTTPKEPAPTTPK---KPAPTTPKEPAPTTPKEPTPTT 264
181 APDXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 240
265 PKE--PAPTTKEP-----APTTPKEPAPTAPK---KPAPTTPKEPAPTTPKEPAPT 311
241 APDXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 300
312 TKEPSTTPPKPAPT--TKSAPTTPKEPAPT-----TKSAPTTPKEPSTPTTK---E 360
301 APDXRP--XPGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 339
361 PAPTPPKPAPTTPK---KPAPTTPKEPAPTTPK---EPAPTTPKAPPTAPKEPAPT 412

```

```

Db 360 SAPDXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 419
Oy 413 TPKETAP-----TTPKILPTTPPEKILAPTTPEKAPPTPEELAPTTPEPTPTTPPEAPPT 468
Db 420 SAPDXRPXGSTAPXAHGVT-----APDXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 471
Oy 469 TPKA-----AAPPTPKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 523
Db 472 APXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVT 530
Oy 524 TTP-----KKPAPELAPTTTKEPTSTTSKRAP--TTPKGTAPPTTPKEPAPTTP 571
Db 531 TAPXAHGVTSAPODXRPXGSTAP--XAHGVTSAPODXRPXGSTAPXAHGVT-----APDXR 585
Oy 572 KEAPPTPKG-----TAPTTKEPAPTTP-----KKPAPELAPTTTKEPTSTTSKRAP 617
Db 586 PXPSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAP--XAHGVTSAPODX 644
Oy 618 KPAP--TTPK-----ETAPTTKEPAPTTP-----KKPAPTTPPETTPPTTSEVSTP 661
Db 645 PXPSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAP--TAPXAHGVTSAPODX 702
Oy 662 TTPKEPT-----IHKSPESTPELSAEPTRKALEMSPEKPEPVPTTKTPAA----- 707
Db 703 DXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODXRPXGSTAPXAHGVTSAPODX 762
Oy 708 -TKPEMTTANKD---TTERDLRTTPETTTAAAPKMTKETATTEKTESKITTATTQVTS 763
Db 763 DXRPXGSTAPXAHGVTSAPODXRPXGSTAP--APXAHGVTSAPODXRPXGSTAPXAHGVTSAPODX 820
Oy 764 TTQDTTPFKITLITLITLAKVTTTKTITTTIIMKPEETAKPKRAINSKATTPKPO 823
Db 821 APDXRPXG-----STAPXAHGVT-----SAPDXRPXGSTAPXAHGVTSAPODXRPXG 868
Oy 824 KPTAPKPKPTTKKPKTPRVKPKETPTTPKPKMTSPELNP--SRTAEMLQTTTPNQ 882
Db 869 -GSTAPXAHGVTSAPODXRP--XPGSTAPXAHGVTSA--PODXRPXGSTAPXAHGVTSAPODX 924
Oy 883 TP 884
Db 925 RP 926

```

```

RESULT 12
US-09-083-116-5
; Sequence 5, Application US/09083116
; Patient No. 6203795
; GENERAL INFORMATION:
; APPLICANT: CHAMBER, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREIVENT, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; City: Alexandria
; State: Virginia
; Country: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,116
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/479,537

```

```

FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 01753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note= "the amino acids spanning
OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-5

```

Query Match 8.2%; Score 423.5; DB 4; Length 1867;
 Best Local Similarity 24.4%; Pred. No. 2.4e-19;
 Matches 235; Conservative 63; Mismatches 461; Indels 203; Gaps 48;

```

75 TSLVANKETVETKTTNKOTSDGKETTSAKETOSIEKTSKD-LAPTSKVLAKPT 133
16 TVLVI-----VTSSGHAASSTPGGEKETSATQSRSSVSSTKKNAVSMSSVLSHS 65
134 P-KAETTKGP--ALTTPKP-----TPPTTKPEASTTPKEPTPTTKAP 176
66 PGGSSSTGQDVTLPATEPAGSAAATWGDVTSVPTRPALGSTTPPAHDVT---SAP 122
177 TTPKEPAPPTT-----KSAPTTKPE-----APTTPKEAPPTTPKEPAPT 217
123 --DNKRPAGSTAPXAHGVTAPDXRPFGSTAPXAHGVTAPDXRPFGSTAPXAHGVT 180

```

```

QY 218 TKE--PAPTTT-----KSAPTTKPEAPPTPK-----KPAPTTKPEAPPTKEPTPTT 264
DB 181 APDXRPAGSTAPXAHGVTAPDXRPFGSTAPXAHGVTAPDXRPFGSTAPXAHGVT 240
QY 265 PKE--PAPTTKEP-----APTTPKEAPPTAK-----KPAPTTKPEAPPTTKEPAPT 311
DB 241 APDXRPAGSTAPXAHGVTAPDXRPFGSTAPXAHGVTAPDXRPFGSTAPXAHGVT 300
QY 312 TKEPPTTPKEPAPT---TKSAPTTTKEPAPT-----TSAPTTTKEPPTT-----E 360
DB 301 APDXRP--XPGSTAPXAHGVTAPDXRPFGSTAPXAHGVTAPDXRPFGSTAPXAHGVT 359
QY 361 PAPTTKPEAPPTPK-----KPAPTTKPEAPPTPK-----EPAPTTTKEPAPTTKEPAPT 412
DB 360 SAPDXRPAGSTAPXAHGVTAPDXRPFGSTAPXAHGVTAPDXRPFGSTAPXAHGVT 419
QY 413 TKEPAP-----TTPKKLTPPTTPEKLAPTTPKEPAPTTPEELAPTTPEEPPTTPEEPAPT 468
DB 420 SAPDXRPAGSTAPXAHGVT---APDXRPFGSTAPXAHGVTAPDXRPX-----PGST 471
QY 469 TPKA-----AAPPTKEPAPPTTPKEPAPTTTPKEPAPTTTKEPAPTTPKGT--APTTLKEPAP 523
DB 472 APXAHGVTAPDXRPAGSTAPXAHGVTAPDXRP--XPGSTAPXAHGVTAPDXRPXGGS 530
QY 524 TTP-----KKPAPKELAPTTTKEPTSTTSKRAP--TTPKGTAPPTPKPEAPPTT 571
DB 531 TAPXAHGVTAPDXRPFGSTAP--XAHGVTAPDXRPXGSTAPXAHGVT---APDXR 585
QY 572 KEPAPTTPKG-----TAPTTLKEPAPTTP-----KKPAPKELAPTTTKEPTSTSD 617
DB 586 PXPSTAPXAHGVTAPDXRPXGSTAPXAHGVTAPDXRPXGSTAP--XAHGVTAPDX 644
QY 618 KPAP--TTPK-----ETAPTTKEPAPTTP-----KKPAPTTPEPTTSEVSTP 661
DB 645 RPXPGSTAPXAHGVTAPDXRPXGSTAPXAHGVTAPDXRPXGGS--TAPXAHGVTAP 702
QY 662 TTPKEPTT-----IHKPDESTEPLSAEPPTPKLENPKRGVPTTKTPA----- 707
DB 703 DKRPXGSTAPXAHGVTAPDXRPXGSTAPXAHGVTAPDXRPXGSTAPXAHGVTAP 762
QY 708 --TPEMTTAKOK-----TTERDLTTPETTTAPAKMTKEATTTETTESKITATTTQVTS 763
DB 763 DKRPXGSTAPXAHGVTAPDXRPXGST--APXAHGVTAPDXRPXGSTAPXAHGVT 820
QY 764 TTTQDTTPFKITLTKTTTLAPRVTTTKKITTETLNNKPEETAKPKDRATNSKATTPKQ 823
DB 821 APDXRPXG-----STAPXAHGVT-----SAPDXRPXGSTAPXAHGVTAPDXRPX 868
QY 824 KPTKAKRKPTSTKPKPTMRVRKPKTTTPRKMTSTMPELNPT--SRIAPAMLOTTRPQ 882
DB 869 -GSTAPXAHGVTAPDXRP--XPGSTAPXAHGVTSA--PDXRPXGSTAPXAHGVTAPDX 924
QY 883 TP 884
DB 925 RP 926

```

RESULT 13
 US-08-479-537A-2
 Sequence 2, Application US/08479537A
 Patent No. 5861381
 GENERAL INFORMATION:
 APPLICANT: CHAMBOU, Pierre
 APPLICANT: KIENY, Marie-Paule
 APPLICANT: LATHE, Richard
 APPLICANT: HAREUVENT, Maira
 TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
 TITILE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
 STREET: P.O. Box 1404
 CITY: Alexandria

STATE: Virginia
COUNTRY: United States
ZIP: 22133-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 - Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y - Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 - Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-08-479-537A-2

Query Match 8.2%; Score 423.5; DB 2; Length 2035;
Best Local Similarity 24.4%; Pred. No. 2.6e-19;

Matches 235; Conservative 63; Mismatches 461; Indels 203; Gaps 48;
QY 75 TSLTVNKETTVERKETTNTKQSTDOKEKXTSAKETOSTEKTSAKD-LAPTSLKVLAKPT 133
| | | | | : | | | | | : | | | | | : | | | | | :
Db 16 TVLTV-----VYSGHASTPGCEKETATONSVPSSSTKNAVAMTSVLSHS 65
QY 134 P-KAETTTKGP--ALTPKRP-----TPTTKEPASTTPKRPPTTTKAP 176
| | | | | : | | | | | : | | | | | : | | | | | :
Db 66 PGSGSSTTGODVTLAPATEPAGSAATWGODVTSVPVTPPALGSTTPPAHDVT---SAP 122
QY 177 TTPKEPAPTTT-----KSAPTPKRP-----APTTTKEPAPTTTKEPAPTT 217
| | | | | : | | | | | : | | | | | : | | | | | :
Db 123 --DNKAPGSTAXAHGVTSAPOXRPXPGSTAPXAHGVTSAPOXRPXPGSTAPXAHGVT 180
QY 218 TKE--PAPTTT-----KSAPTPKRPAPTTPK---KPAPTTKEPAPTTKEPPTT 264
| | | | | : | | | | | : | | | | | : | | | | | :
Db 181 APDXRXPXPGSTAPXAHGVTSAPOXRPXPGSTAPXAHGVTSAPOXRPXPGSTAPXAHGVT 240
QY 265 PKE--PAPTTKEP-----APTTKEPAPTAOK---KPAPTTKEPAPTTKEPAPTT 311
| | | | | : | | | | | : | | | | | : | | | | | :
Db 241 APDXRXPXPGSTAPXAHGVTSAPOXRPXPGSTAPXAHGVTSAPOXRPXPGSTAPXAHGVT 300
QY 312 TKEPSPPTTKEPAPTT--TKSAPTTTKEPAPTT-----TKSAPTTKEPSPPTTTK-----E 360
| | | | | : | | | | | : | | | | | : | | | | | :
Db 301 APDXRP--XPSTAPXAHGVTSAPOXRPXPGSTAPXAHGVTSAPOXRPXPGSTAPXAHGVT 359
QY 361 PAPTTTKEPAPTTPK-----KPAPTTTKEPAPTTPK-----EPAPTTTKEPAPTAOK 412
| | | | | : | | | | | : | | | | | : | | | | | :
Db 360 SAPDXRXPXPGSTAPXAHGVTSAPOXRPXPGSTAPXAHGVTSAPOXRPXPGSTAPXAHGVT 419
QY 413 TPKEPAP---TPPKLTPPTTPEKLAPTTPEKAPPTPEELAPTTPEEPPTTPEEPAPTT 468
| | | | | : | | | | | : | | | | | : | | | | | :
Db 420 SAPDXRXPXPGSTAPXAHGVTSAPOXRPXPGSTAPXAHGVTSAPOXRPXPGSTAPXAHGVT 471
QY 469 TPKA-----AAPNTPKPEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 523
| | | | | : | | | | | : | | | | | : | | | | | :
Db 472 APXAHGVTSAPOXRPXPGSTAPXAHGVTSAPOXRP--XPSTAPXAHGVTSAPOXRPXPGS 530
QY 524 TTP-----KKPAPKELAPTTTKEPSTISOKPAP--TTPKSTAPTTKEPAPTT 571
| | | | | : | | | | | : | | | | | : | | | | | :
Db 531 TAPXAHGVTSAPOXRPXPGSTAP--XAHGVTSAPOXRPXPGSTAPXAHGVTSAPOXRP 585
QY 572 KEPAPTTPKG---TAPTTLKEPAPTTT-----KKPAPKELAPTTTKEPSTISD 617
| | | | | : | | | | | : | | | | | : | | | | | :
Db 586 PXXGSTAPXAHGVTSAPOXRPXPGSTAPXAHGVTSAPOXRPXPGSTAP--XAHGVTSAPOX 644
QY 618 KPAP--TTPK---ETAPTTTKEPAPTTT-----KKPAPTTTKEPAPTTTSEVSTP 661
| | | | | : | | | | | : | | | | | : | | | | | :
Db 645 RPXPGSTAPXAHGVTSAPOXRPXPGSTAPXAHGVTSAPOXRPXPGS--TAPXAHGVTSA 702
QY 662 TTTKEPPTT-----THKSPDESTPELSAERTPKALENSKKEGCVPTTKTPAA----- 707
| | | | | : | | | | | : | | | | | : | | | | | :
Db 703 DXRXPXGSTAPXAHGVTSAPOXRPXPGSTAPXAHGVTSAPOXRPXPGSTAPXAHGVTSA 762
QY 708 -TKPEMTTAKDK---TTERDLRTTPETTTAAKMTKETATTEKTTESKTTATTTQVTS 763
| | | | | : | | | | | : | | | | | : | | | | | :
Db 763 DXRXPXGSTAPXAHGVTSAPOXRPXPGST--APXAHGVTSAPOXRPXPGSTAPXAHGVT 820
QY 764 TTTQDTTPRKITLTKTTTAPKVTTKKTTTTELNKKPEETAKKPRDRTATNSKATTPK 823
| | | | | : | | | | | : | | | | | : | | | | | :
Db 821 APDXRXPXPG-----STAPXAHGVT-----SAPDXRXPXPGSTAPXAHGVTSAPOX 868
QY 824 KPAPKAPKKTSTKPKPTMPVRKPKTTTPRKMTSTMPKELNPT--SLIAEMLQDTTRPNQ 882
| | | | | : | | | | | : | | | | | : | | | | | :
Db 869 -GSTAPXAHGVTSAPOXRP--XPSTAPXAHGVTSA--POXRPXPGSTAPXAHGVTSAPOX 924
QY 883 TP 884
| | | | | : | | | | | : | | | | | : | | | | | :
Db 925 RP 926

RESULT 14
US-09-083-116-2

Sequence 2, Application US/09083116
Patent No. 6203795
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREIVENT, Maira
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,116
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide

LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-2
Query Match 8.2%; Score 423.5; DB 4; Length 2035;
Best Local Similarity 24.4%; Pred. No. 2,6e-19;
Matches 235; Conservative 63; Mismatches 461; Indels 203; Gaps 48;
QY 75 TSLTVNKTETVETKETTINKQSTDGKREKTSKAKETOSTEKSAD-LAPTSKVLAKPT 133
DB 16 TVLTV-----VTSCHASTPGCEKETSTQKSSVSSSTKMAVSMTSVLSHS 65
QY 134 P-KAETTKGP--ALTTPKEP-----PTTPKEPASTPKPEPTTIKAP 176
DB 66 PGSGSSTTGQDVTLLADPATEPASGSAATWQDVTVPVTPALGSTPPAHDT--SAP 122
QY 177 TPKEPAPTT-----KSAPTPKEP-----APTTPKEPAPTTKEPAPT 217
DB 123 --DNKRAPGSTAPXAHGVTAPDXRXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVT 180
QY 218 TKE--PAPTT-----KSAPTPKEPAPTTKEP-----KPAPTPKEPAPTTKEPPTT 264
DB 181 APDXRXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVT 240
QY 265 PKE--PAPTTKEP-----APTTPKEPAPTPAR-----KPAPTPKEPAPTTKEPAPT 311
DB 241 APDXRXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVT 300
QY 312 TKEPSTTPKEPAPTT--TKSAPTTTKEPAPT-----TKSAPTTKEPSPPTTK--E 360
DB 301 APDXRXP--XPGSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVT 359
QY 361 PAPTPKEPAPTPK-----KPAPTPKEPAPTTK-----EPAPTTTKKRAPAKPEAPT 412
DB 360 SAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVT 419
QY 413 TPKEPTAP-----TPKKLLPTTPEKLAPTTPKEPAPTTPEELAPTTPEPTTPEEAPT 468
DB 420 SAPDXRXPSTAPXAHGVT-----APDXRXPSTAPXAHGVTAPDXRXP--PGST 471
QY 469 TPKA---AAPTTPKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTT 523
DB 472 APXAHGVTAPDXRXPSTAPXAHGVTAPDXRXP--XPGSTAPXAHGVTAPDXRXP 530
QY 524 TTP-----KKRAPKELAPTTTKEPTSTSDKRAP--TPPKGAPTPKEPAPT 571
DB 531 TAPXAHGVTAPDXRXPSTAP--XAHGVTAPDXRXPSTAPXAHGVT-----ABDXR 585
QY 572 KEPAPTTPKG---TAATTTKEPAPTT-----KKRAPKELAPTTTKEPTSTSD 617
DB 586 PXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAP--XAHGVTAPDX 644
QY 618 KRAP--TTPK---ETAPTPKEPAPTT-----KKRAPTTPEPTTSEVSTP 661
DB 645 RPXGSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPXGS--TAPXAHGVTAP 702
QY 662 TTTKEPTT-----IKSPDESTPELSAETTPALNSPREPVPTTKPAA----- 707
DB 703 DXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAPDXRXPSTAPXAHGVTAP 762
QY 708 -TRPEMTTADK---TTERDLRTTPEPTTAAPKMTEKETAATTEKTSKITTATTTGVS 763
DB 763 DXRXPSTAPXAHGVTAPDXRXPST--APXAHGVTAPDXRXPSTAPXAHGVT 820
QY 764 TTTODTTPFKITTLKTTLLAPKVTTTTKKITTTEIMNKPEETAPKPDRAINSKATTPKQ 823

Mon Apr 29 08:35:34 2002

us-09-556-246-1_copy_200_1167.ra

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:19:22 ; Search time 82.85 seconds
(without alignments)
931.380 Million cell updates/sec

Title: us-09-556-246-1_copy_200_1212

Perfect score: 5416
Sequence: 1 VKDNKKNRKKRPKPPV.....SPIDVTTCRCNCEGKFFFK 1013

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1062.5	19.6	3020	2	A43932
2	942	17.4	1664	2	T18262
3	853	15.7	1489	2	T31108
4	802	14.8	1274	2	T16251
5	789.5	14.6	1367	1	S48478
6	784	14.5	1188	2	S49915
7	776	14.3	2187	2	T30826
8	772	12.4	1344	1	A35175
9	671	12.4	1151	2	T18535
10	659.5	12.2	1229	2	T25697
11	633	11.7	489	2	T11622
12	632.5	11.7	7962	2	T18346
13	632	11.7	761	2	C84672
14	629	11.6	3507	2	T34513
15	626.5	11.6	990	2	T51618
16	625	11.5	6642	2	T29757
17	623.5	11.5	839	2	T75518
18	622	11.5	3570	2	T45025
19	620.5	11.5	971	2	T19431
20	607.5	11.2	801	2	T29018
21	607.5	11.2	924	2	S27923
22	605	11.2	379	2	S50125
23	559.5	10.3	350	2	S22456
24	556.5	10.3	856	2	T16543
25	556.5	10.3	875	2	S23760
26	555.5	10.3	1630	2	A53577
27	553	10.2	2232	2	T34434
28	551	10.2	620	2	S06733
29	545	10.1	873	2	A47283

30	542	10.0	369	2	S20500	hydroxyproline-ric
31	538.5	9.9	416	2	TJ0465	extensin precursor
32	533	9.8	1087	1	QPM5H	neurofilament trip
33	532	9.8	756	2	T27642	hypothetical prote
34	530.5	9.8	1162	2	TJ0557	exo-alpha-sialidas
35	530	9.8	865	2	A47282	calcium-binding pr
36	522	9.6	328	2	TJ00985	hydroxyproline-ric
37	522	9.6	1459	2	T33271	hypothetical prote
38	518	9.6	813	2	S70795	vsda protein precu
39	518	9.6	866	2	T45462	membrane glycoprot
40	518	9.6	1072	1	A37221	neurofilament trip
41	512.5	9.5	867	2	T45463	membrane glycoprot
42	509.5	9.4	662	2	A45155	mucin FIM-C.1 - Af
43	508	9.4	1832	2	T31113	mucin-like glycopr
44	505	9.3	700	2	A54641	interspersed repea
45	504.5	9.3	606	2	A43427	neurofilament trip

ALIGNMENTS

RESULT 1

A43932

mucin 2 precursor, intestinal - human (fragments)

N.Alternate names: mucin SMUC-41

C.Species: Homo sapiens (man)

C.Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 05-Nov-1999

C.Accession: A49963; A45106; B45106; A43932; B33532; A61257; F00328; P00329

R.Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.

J. Biol. Chem. 269, 2440-2446, 1994

A.Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t

A.Reference number: A49963; MUID:94132002

A.Accession: A49963

A.Molecule type: mRNA

A.Residues: 1-639 <GU1>

A.Cross-references: GB:L21998

R.Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.

J. Biol. Chem. 267, 21375-21383, 1992

A.Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up

A.Reference number: A45106; MUID:93016075

A.Accession: A45106

A>Status: not compared with conceptual translation

A.Molecule type: mRNA

A.Residues: 626-1895 <GU2>

A.Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396

A.Note: sequence extracted from NCBI backbone (NCBI:116706)

A.Accession: B45106

A>Status: not compared with conceptual translation

A.Molecule type: mRNA

A.Residues: 2037-3020 <GU3>

A.Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398

A.Note: sequence extracted from NCBI backbone (NCBI:116638)

R.Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, J. Clin. Invest. 88, 1005-1013, 1991

A.Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym

A.Reference number: A43932; MUID:91358717

A.Accession: A43932

A.Molecule type: DNA

A.Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>

A.Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864

A.Note: sequence inconsistent with the nucleotide translation

A.Note: sequence extracted from NCBI backbone (NCBI:55749; NCBI:55750)

R.Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.

J. Biol. Chem. 264, 6480-6487, 1989

A.Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evl

A.Reference number: A33532; MUID:89197956

A.Accession: B33532

A.Molecule type: mRNA

A.Residues: 1916-2193 <GU4>

A.Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874

A.Experimental source: Intestine

R.Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

J. Clin. Invest. 87, 77-82, 1991
A:Title: Human bronchus and intestine express the same mucin gene.
A:Reference number: A61257; MUID:91086481
A:Accession: A61257
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 'T',1925-1948,'TTS',1952-1954 <JAN>
A:Experimental source: bronchus
R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-t
A:Reference number: P00328; MUID:92198477
A:Accession: P00328
A:Molecule type: mRNA
A:Residues: 2328-2468 <XUG>
A:Cross-references: GB:M6523
A:Experimental source: small intestine
A:Accession: P00329
A:Molecule type: protein
A:Residues: 2328-2342,'K',2344-2354 <XUG1>
C:Genetics:
A:Gene: GDB:MUC2
A:Cross-references: GDB:120203; OMIM:158370
A:Map position: 11p15.5-11p15.5
C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von
C:Keywords: glycoprotein; intestine; tandem repeat
A:Z166-2834/Domain: von Willebrand factor type C repeat homology <WMC>

[illegible][illegible]

```

RESULT      2
T18262
S-layer protein - Clostridium thermocellum
C.Species: Clostridium thermocellum
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C.Update: T18262
R.Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A.Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e.
A.Reference number: Z18847; MUID:93209931
A.Accession: T18262
A.Status: preliminary; translated from GR/EMBL/DDBJ
A.Molecule type: DNA
A.Residues: 1-1664 <FUJ>
A.Cross-references: EMBL:X67506; MID:g296879; PID:g296881; PIDN:CAA47841.1

Query Match          17.4%; Score 942; DB 2; length 1664;
Best local similarity 33.0%; Pred. No. 1..36-33;
Matches 276; Conservative 83; Mismatches 312; Indels 166; Gaps 43;

QY      128 VLAKPTP-KAEETTKGPAITTPKEPTTPPKESPATSKPEPTPTTIKSAPITPKKEPATT 186
         |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB       758 VIQAPAIKASDESDIPDTSPDEPTPS-----DEPTPS--DEPTSPDEPTPS 804
         |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      187 TKSAPTTPKEPAPTTPKKEPAPTTKEPA---PTTPKEPAPTTPKSAPTTPKEPAPTTK 242
         |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      805 EPTPSETPEEPPIPTDTPSEPTSPDEPTPSDEPTPSDEPTPSDEPTPEEPPIPDTP 864
         |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY     243 KPAPTTPKEPAPTTPKKEPAPTTPKKEPAPTTPKEPAPTTPKKEPAPTTPKKEPAP 301
         |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


Db 865 SDEPTSPDEPTSS--DEPTPS--DEPTP--SDEPTSPSEPTPEPIPTDTPSDEPTSPDEPTP 919
QY 302 TTPKEPA--PTTKESPPT--TPKEPAPTTKS--APTTPKEPAPTTKSAPTTPKESSPTT 357
Db 920 SDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 11
QY 358 TKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT--TPKEPAPTTPKAPPAKPAKPAKPAKPAK 416
Db 969 SDEPTSS--DEPTPS--DEPTPSDEPTSPSEPTPEPIPTDTPSDEPTSPDEPTPSD---- 1019
QY 417 TAPTTKATLPTTPKEKAPPTTPKEKAPPTPEELAPTTPEEPPT--TPKEPAPTTPKAAAP 475
Db 1020 -----EPPTSPD--EPPTSPD-----EPPTSPDEPTSPSEPTPEPIPTDTPSDEPT 1064
QY 476 TPKKEPAPTTPKEPA--PTTPKEPAPT--TPKEPAPTTPKGTATLKEKAPPTTPKPAK 532
Db 1065 TSPDEPTSPDEPTSPDEPTSPDEPTSPSEPTPEPIPTDTPSDEPTSPDEPTPS--DEPTPS 1122
QY 533 ELAPTTTKAPT--STSDKAPATTPKGTATTPKEPAPTTPKEPAPTTPKGTATLKEPA 591
Db 1123 D-EPPTSPDEPTSPSEPTPEPIPTDTPSDEPTSPDEPTPS--DEPTPS--DEPTSPDEPT 1175
QY 592 PT--TPKPAKAPKELAPTTGPTSTSDKAPT--TPKTAATTPKEPAPT--TPKKAAPT 646
Db 1176 PSEPTPEP-----IPTDTPSDEPTSPDEPTSPDEPTSPD--EPPTSPDEPTSPSEPTPEPIPT 1229
QY 647 TPEPTPTTSEVSTPTTKPTTTHKSPDESTPELSAEPTRPALENSPKPEPGVPTTKPA 706
Db 1230 DTPSPDEPTSPD--EPPTSPDEPT-----PSDEPTP--SDEPTP--SEPTPEP--IPTDTPSD 1277
QY 707 ATKPEPTTAKKKTTERDLITTPETTTAAKMKETATTEKTESKITATTTQVSTTT 766
Db 1278 EPPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPSEPTPEPIPTDTPSDEPTPS 1337
QY 767 OOTPEKITTLLTTTLARVTTTKTITTTETIMNKPEELAKKDAATNSKATTPKQ--KP 825
Db 1338 DEPTSPDEPT-----PSDEPTSPDEPTSPDEPTSPDEPTP 1372
QY 826 TPAKPKPTSTKPKTMPRVRKPTTPTPKMTS-----TMPELNTSRJA 870
Db 1373 SEPTPEPTPTTP-----TPPTSTPTSSGSGSGSGSGGGGGGTVPTSPPTPTS--- 1424
QY 871 EMLQTTTPRPNQTPNSKLEVNPKSEDAAGAGETPHMLLRPHVMEVTPMDYLP 927
Db 1425 -----KPTSTPAPTLEI--EPPTSDVPGAIGENRAVLRGY-----PDGSPRP 1465

RESULT 3
T31108
cyst germination specific acidic repeat protein precursor - Phytophthora infestans
C:Species: Phytophthora infestans (potato late blight agent)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31108
R:Goerhardt, B.
submitted to the EMBL Data Library, April 1998
A:Reference number: Z20986
A:Accession: T31108
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1489 <GOE>
A:Cross-references: EMBL:AF061185; NID:93851513; PID:93851514; PIDN:AACT2308.1
C:Genetics: car90
A:Gene: car90

Query Match 15.7%: score 853; DB 2: length 1489;
Best Local Similarity 32.3%: Pred. No. 7.6e-30;
Matches 355; Conservative 46; Mismatches 475; Indels 222; Gaps 49;

QY 9 TTKKPTPRPVVDAGSGIDNGCFVTPPTDSTT-----OHNKVSTSPKITTAKKPIPR 62
Db 338 TPVAPTEKRYDVEETTYTTESTYAPTAKSETNAPTERNHVHIEKRCDEVTAKAPTEET 397

QY 63 PSLP-----PNSDT---SKETSLVNNKETTYETKETT---TNKQSTND 100
Db 398 TVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 457
QY 101 GKEKTTSAKETOSIEKTSKADLAPTSKVLAKPTPAEETTKCPALUTPKREPTTPKKEA 160
Db 458 TEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 514
QY 161 STTPKEPTPTTKSAPTTPKEPAPT--TKSAPTTPKEPAPT--TPKEPAPTTPKAPAPT 217
Db 515 EETTPYEPTPEET--TVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 573
QY 218 TKEPAPTTPKSAPT-----TPKEPAPTTPKKAAPTTPKEPAPTTPKKEPTTPKE 267
Db 574 TVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTV 632
QY 268 P-----AP--TPKEPAPTTPKKEP-----APAPKPKAPAPTTPKEPAPTTP 304
Db 633 PTEETTVASTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTE 692
QY 305 KEPAAPT--TPKPSPTTPKEP-----APT--TKSAPT-----TPKEP-----A 339
Db 693 TVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTV 752
QY 340 PT--TKKSAPT--TPKEPSPTTPKEP-----APTTPKEPAPTTPKKAAPTTPKEPAPT-- 388
Db 753 PTEATTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTE 812
QY 389 TPKEPAPTTPKKAAPTTPKEPAPTTPKETAAPTTPKKTLPPTPEKILAPPTTPKEPAPTTPEE 448
Db 813 TPYEPTTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTV 870
QY 449 LAPT--TEETPTPTPEEP-----APT-----TPKAAAPTTPKEPAPTTPKEP- 489
Db 871 YAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 930
QY 490 -----APTTPKEPAPTTPKETAAPTTPKGTAPT--LTKKAPAPTTPKKAAPT--ELAPT--TT 539
Db 931 EETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTV 990
QY 540 KEPTSTSDKAPATTPKGTAPPTTPKEPAPTTPKEP-----APTTPKGTAPT----- 585
Db 991 YAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 1050
QY 586 -----LTKKAPAPTTPKAP-----APKE--LAPT--TPKGTSTSDKAPAPTTP 624
Db 1051 EETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTV 1110
QY 625 KETAPPTTPKEPAPTTPKKAAPTTPPTPEPTSPSTPTTKAPT--TINKSPDESTPELS 682
Db 1111 YAPAEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 1169
QY 683 AEPTPKA--LENSPKEPG-----VPTTKPAATKDEMTTAKDKTTERDLRPTTP-----ETT 732
Db 1170 TEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTE 1229
QY 733 TPAKMKETATTPTEKTESKITATTTQVSTTQDIT-----PEKITT--LKT 779
Db 1230 TVAP-----TEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 1284
QY 780 TTLAPKVTTTKKITTTPETIMNKPEETAKPKDRAT--NSKATTPRP--OKPTAKPKAPT-- 833
Db 1285 TTVAPTEATTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAP 1344
QY 834 -----STKRP--KTPRVRKPKTTPTPKMTSTMBELNPTSRIAEMLQTTTPR--N 881
Db 1345 PAESTSTVSTEKPCONTLEEFDEPTDEPTD--PSDEPTDEPTDEPTDLPDDEPTSCDN 1402
QY 882 QTPNSKLEVNPKSEDAAG 899
Db 1403 OGINGIGVENKVRYNAG 1420

RESULT 4
 T16251
 Hypothetical protein F35A5.1 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
 C:Accession: T16251
 R:Leimbach, D.
 submitted to the EMBL Data Library, January 1996
 A:Description: The sequence of *C. elegans* cosmid F35A5.
 A:Reference number: 218485
 A:Accession: T16251
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1274 <LEI>
 A:Cross-references: EMBL:046675; NID:g1166613; PID:g1166621; PIDN:AAH52641.1; GSPDB:GN00
 A:Experimental source: strain Bristol N2; clone F35A5
 C:Genetics:
 A:Gene: CESP:F35A5.1
 A:Map position: X
 A:introns: 1272/2

Query Match 14.8%; Score 802; DB 2; Length 1274;
 Best Local Similarity 28.0%; Pred. No. 1e-27;
 Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;

OY 6 KNRTRK-KP-----TP-----KPVVDEAGSLDNGD-----FKVTPPTSTQHNKVSST 50
 DB 274 KNTTKKMKPWEDETFVEEVKEBPVEKAPVLEKKDPAPAAKADSPSKAAPKVEPS 333
 OY 51 PKTTAKPI-NPRPSLPNSDTSKETSILVNNKETVETETETTNKOTSTDCKEKTTSAK 109
 DB 334 SPVVPPTPVKNPKVKKKPPMEVDEDAEAEVKKPSAEKPTVLAKRKEPSSSTPSDDS 393
 OY 110 ETOSIEKTSADLAPTSTVLAKEPTPKAE-----TTTKGA-----LTPP 148
 DB 394 PKAAADVAVPRDSSPKKAPPLQADPKAEVPPPVKNPKVKKYKPPWEVDEDEDFVEEVKQ 453
 OY 149 KEPTPTTP-----KEPASTTPKEP-----TP-TTKSAPTPKKEPAPTTSAPTTPKEPAP 199
 DB 454 EAAKAKTTPVLAKRKEPAKDTAKPATSTPTTPPKKDPVKRDPSSPKVAAKPPSAQAPA 512
 OY 200 TTTKEPA-----PTPKKEPAPTTPKEPAPT-----TTKSAPTTP 233
 DB 513 TPVKNVVKMKRPPWEDETFPADVSKPTLAKKTPSLAKKDPAPAKSLKAKDTKAPAKP 572
 OY 234 KEP-----APTTKKKAPPTPK-----PAPTTPKEPTPTTPPKKEPAPTTPKEPAPTTPK-EP 283
 DB 573 RDPSPKKVAVPTAPKEKTPVLAKKEPAGADSKTEPEKSKRDPSPKAVPAKPVKTEV 632
 OY 284 APAPKPPAP-----TTKEPAPTTPKEPAPTTPKEP-----SPTTPKEPAPTTPKSAPT 333
 DB 633 APAVAVKKEPISKPKDTAPKAPKAPNSVVP-PVPVKNPVKMKRPPWEDDAPAKPVSLPE 691
 OY 334 TTKEPAPTTPKSAPTTP-KEPSPPTTKKEPAPTTPK--EPAPTTPKKAP--TTPKKEPAPT 388
 DB 692 PEKK-TPVLAKKAPTKPDESAADPVSGPSSKDKLAKKAPVPRDPSPKAVPIKAPK 750
 OY 389 TPKEPAPTTPKKAPPA-PKEPAPTTPKEPAPTTPPKLTPPTTP-----EKLAPTTTPKPA 442
 DB 751 T--EVPAVAVKKEPAPKSDPSKRAK--AEPNSP--VPEPTPVKNPVKMKRPPWEDDA 805
 OY 443 PTPPEELAPTTPEEP-----TPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTP--PK 496
 DB 806 PABEVNVPDEEKTPVLAKKTPVKRDPSPKAVPAKPSIKTADAPVSVKKEPEVSKPE 865
 OY 497 PAPTTPKEPAPTTPKGAAPTTLKEPAPTTPKKAPKELAPT--TTKEPTSTTSKAPAPTTP 555
 DB 866 PSPKAPNSPVVP-----PVPVKNPVKMK--KPPWEDEDETFVEEVKQSE--PEKKTPEVLA 918
 OY 556 KGTAPTTPKKEPAPTTPKEPAPTTPKGAAPTTLKEPAPT-----TPKKAPAP-----KEL 603
 DB 919 K-KEPKPKD-APKVAAPKDPSPKAVP--KEPAVAAKPRDLSPKALPIPIANTQEA 974

OY 604 APTTKKCPST-----TNSDKP-----APTTPKET-----APTTPKEPAPTTPKEPAPT 646
 DB 975 PTPVKNPVKMKRPPWEDEDEPAPVSAPEPEKPTPVLAKAAPKAPRPP-----SPKKAAPV 1031
 OY 647 TPETPTPTTSEVSPPTTTPKEPTTIKSP-----DESTPELSA-EP--TPKALENSPKKEG 698
 DB 1032 AAK-PDPKRIPEP-PPTPVKNPVKMKRPPWEDEDEPSEVSAPEPEKPTPVLAKAAPKAPA 1089
 OY 699 V-----PTTKTPAAT-----KPEMTTAAKDKTTERLRTPTPT--TAAPK 737
 DB 1090 TKPDESAADPVSGTSPKDPKLSKAPAVEKPPPTDPKDDKIKESPAKPEKAPPAAPK 1149
 OY 738 MKKETATTTKETESKITATTTQVSTTQDTTPPKITTLKTTTLAPKVTTKKTTTTE 797
 DB 1150 KKKPWWDDPDEPDEDFYPAKSKKPTIEDPADPLG-----GPKTKDPK----- 1193
 OY 798 INKKEETAKPKDRATNSKATTPKPKOKPTKADKPTSKKPKTMVRVKPK----- 848
 DB 1194 -LNKKAPAKKPTK-----PKKEVSKKEPKKPTKEPKP--AADKKMKPWEDEDDPEPE 1243
 OY 849 ---TTPPKKMTSTMP 861
 DB 1244 ADETPAPKRPDTEDP 1259

RESULT 5
 S48478
 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR01
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
 C:Accession: S48478; A26877; B26877; S27281; J06123
 R:Rowley, K.
 submitted to the EMBL Data Library, October 1994
 A:Reference number: S48478
 A:Accession: S48478
 A:Molecule type: DNA
 A:Residues: 1-1367 <RDN>
 A:Cross-references: GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GSPDB:GN00009;
 J.Yamashita, I.; Nakamura, M.; Fukui, S.
 J. Bacteriol. 169, 2142-2149, 1987
 A:Title: Gene fusion is a possible mechanism underlying the evolution of STAL.
 A:Reference number: A91831; MUID:87194600
 A:Accession: A26877
 A:Molecule type: DNA
 A:Residues: 1-242 <YAM>
 A:Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
 A:Accession: B26877
 A:Molecule type: DNA
 A:Residues: 762-1331 <YA2>
 A:Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
 A:R:Paro, J.M.; Ianez, E.; Zalacain, M.; Clares, M.G.; Jimenez, A.
 FEBS Lett. 239, 179-184, 1988
 A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from *Sacc*
 A:Reference number: S27281; MUID:89031230
 A:Accession: S27281
 A:Molecule type: DNA
 A:Residues: 1-31 <PAR>
 A:Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
 R:Lambrecht, M.G.; Bauer, F.F.; Marmur, J.; Preorius, I.S.
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
 A:Title: Muc1, a mucin-like protein that is regulated by Msi10, is critical for pseud
 A:Reference number: J06123; MUID:96323237
 A:Accession: J06123
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-1367 <LAM>
 A:Cross-references: GB:U30626; NID:g1304386; PIDN:AA49609.1; PID:g1304387
 C:Genetics:
 A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
 A:Cross-references: MIPS:YIR019C; SGD:S0001458
 A:Map position: 9R

A:Note: GenBank entry HUMMUCAB includes one copy of the tandemly repeated sequence
R:lan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
J. Biol. Chem. 265, 15294-15299, 1990
A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
A:Reference number: A35887; MUID:90368716
A:Accession: A35887
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19, 29-1109, 'S', 1111-1339, 'A', 1341-1344 <LAN>
A:Cross-references: GB:J05582; NID:q189598; PIDN:AAA60019.1; PID:q189599
A:Note: Genbank entry HUMPMNU contains four fewer copies of the tandemly repeated sequence.
R:Weschner, D.H.; Harevel, M.; Teslaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.
Eur. J. Biochem. 189, 463-473, 1990
A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may generate
A:Reference number: S10571; MUID:90276413
A:Accession: S10572
A:Molecule type: mRNA
A:Residues: 1-19, 29-155, 'P', 157-175, 'P', 177-182, 'A', 184-212, 1033-1344 <WRE>
A:Cross-references: EMBL:X52229; NID:q37053
R:Weschner, D.H.
submitted to the EMBL Data Library, March 1990
A:Reference number: S40293
A:Accession: S40293
A:Molecule type: mRNA
A:Residues: 1-19, 29-155, 'P', 157-175, 'P', 177-182, 'A', 184-212, 1033-1037, 'A', 1039-1344 <WR2>
A:Cross-references: EMBL:X52229; NID:q37053; PIDN:CAA6478.1; PID:q37054
R:De, M.; Siddiqui, J.; Kufe, D.
Biochem. Biophys. Res. Commun. 165, 644-649, 1989
A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
A:Reference number: A36735; MUID:90088473
A:Accession: A36735
A:Molecule type: mRNA
A:Residues: 1-142, 'Q', 144-162, 'Q', 164-168 <ABE>
A:Cross-references: EMBL:X31833; NID:q181543; PIDN:AAA3757.1; PID:q181543
R:Masuzawa, Y.; Miyachi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.
J. Biochem. 112, 609-615, 1992
A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu
A:Reference number: JX0235; MUID:93123189
A:Accession: JX0066
A:Molecule type: mRNA
A:Residues: 998-1011, 'ES', 1014-1017, 1018-1032, 'T', 1034-1037, 1038-1057 <MAS>
A:Experimental source: gastric carcinoma cell
R:Zrlhan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Weschner, D.H.
FEBS Lett. 356, 130-136, 1994
A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine
A:Reference number: S31026; MUID:95080414
A:Accession: S31026
A:Contents: annotation
A:Note: Underlined tyrosine residues in the carboxyl-terminal non-repetitive region at
C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
partial repeats. The repeat shown is defined by SmaI nuclease sites.
C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
C:Gene: GDB:MUC1; PUM
A:Cross-references: GDB:120705; OMIM:158340
A:Map position: 1q21-1q23
A:Introns: 20/1: 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
C:Superfamily: polymorphic epithelial mucin
C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphis
E:1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>
E:1-62/Region: mucin 1 amino-terminal non-repetitive
E:1-23/Domin: signal sequence #link PREA #status predicted <SIGA>
F:1-19, 29-33/Domin: signal sequence #link PREB #status predicted <SIGB>
F:1-19, 29-33/Domin: signal sequence #link PREB #status predicted <SIGB>
F:1-19, 29-212, 1033-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>
F:1-13-1017/Region: 20-residue repeats (GSTAPRANGVTSAPDTRAP) #status predicted <SIGA>
F:1-13-1344/Region: mucin 1 carboxyl-terminal non-repetitive
F:1-1245-1272/Domin: transmembrane #status predicted <TRM>
F:1-1046, 1064, 1118, 1144, 1222/Binding site: carbohydrate (Asn) (covenant) #status predicted
F:1213/Binding site: phosphate (Tyr) (covenant) #status predicted

Query Match

12.4% score 672; DB 1; Length 1344;

Best Local Similarity 27.8%; Pred. No. 4, 1e-22; Matches 310; Conservative 78; Mismatches 472; Indels 256; Gaps 56;

```
77 LVNKEETVEETKEETTTNK--OTSDGKEKTTSAKTESIETKSAKD-LAPTSKVLAKPT 133
15 LVLTATATAPKRAIVTGTGGHASTGTGGEKETSATDRSSVPSSTENAVSWTSVLSHS 74
134 P-KAETTTGCP-ALTPKEP-----TPTPKEBASTP-----KE 166
75 PGSGSSTTGODVTLAPATEPASGAATWGOVTSVTPRPALGSTTPRANHYTSAPDNK 134
167 PPTPT-----IKSAPTPKEBAPPTT-----KSAPTTKEBPPTTKER-----A 206
135 PARGSTAPRANGVTSAPDTR--RPAGSTAPRANGVTSAPDTR--RPAGSTAPRANGVTS 190
207 PTPKEBAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTKB 252
191 PDT--RPAGSTAPRANGVT--SAPDTRAPAGSTAPRANGVTSAPDTRAPAGSTAPRANG 247
253 --APTPKEPTPTTP-----KEBAPTTKEP-----APTPKEBAPTPAK---- 289
248 TSAPDTRAPAGSTAPRANGVTSAPDTRAPAGSTAPRANGVTSAPDTRAPAGSTAPRANG 307
290 KPAPTPKBPAPTPPK-----EPAPTTTKEPPTTPKEBAPTT--TKSAPTTKEBAPTT 342
308 TSAPDTRAPAGSTAPRANGVTSAPDTRAPAGSTAPRANGVTSAPDTRAPAGSTAPRANG 367
343 TYSAPTPKEPSPPTTKER-----APTPKEBAPTPK-----KPAPTPKBPAPTPK- 391
368 T-SAPDT--RPAGSTAPRANGVTSAPDTRAPAGSTAPRANGVTSAPDTRAPAGSTAPRA 424
392 ---EPAPTTTKBAPTPKBPAPTPKBPAPTPKBPAPTPKBPAPTPKBPAPTPKBPAP 447
425 HGVTAPDTRAPAGSTAPRANGVTSAPDTRAPAGSTAPRANGVTSAPDTRAPAGSTAPRA 484
448 ---ELAPTTKEBPTPTTP-----EAPAP--TPKA---AAPTTPKEBAPTPK- 487
485 HGVTAPDTRAPAGSTAPRANGVTSAPDTRAPAGSTAPRANGVTSAPDTRAPAGSTAPRA 544
488 ---EPAPTTKBPAPTPPKETPTTPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPK 534
545 HGVTAPDTRAPAGSTAP--RANGVTSAPDTRAPAPSTAPRANGVTSAPDTRAPAGST 600
535 APTTKKEPTSTSDKPAP--TPPKSTAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBP 588
601 AP-RANGVTSAPDTRAPAGSTAPRANGVTS-----APDTRAPAGSTAPRANGVTSAPDTR 655
588 EPAPTPP-----KKAPKEIAPPTTKGPTSTSDKPAP--TPPK---ETAPTPP 632
656 AFGSTAPRANGVTSAPDTRAPAGSTAP--RANGVTSAPDTRAPAGSTAPRANGVTSAPDTR 714
633 KEBAPTPP-----KKRAPTPETTPPTTSVSTPTTKETP-----IKNSPD 675
715 PARGSTAPRANGVTSAPDTRAPAGS--TAPRANGVTSAPDTRAPAGSTAPRANGVTSAPD 772
676 ESTPELSAEPTRKALENSPKERGVPPTTKTPAA-----TKPEMTTAKDK--TTARD 724
773 TRPAGSTAPRANGVTSAPDTRAPAGSTAPRANGVTSAPDTRAPAGSTAPRANGVTSAPD 832
725 LKTPETP-----TAAPKATKETATTEKTESKITATTTGVTSTTTODTPP--FKIT 775
833 TRPAGSTAPRANGVTSAP-----DTRAPAGSTAPRANGVTSAPDTRAPAGSTAPRANGVT 888
776 TLKTTTLAPKVTYTTTKTTT--EIMNKEETAPKPDATNSKATTPRPOKPTAPKPKPT 833
889 SAPDTRAPAGSTAPRANGVTSAPDTRAPAGSTAPRANGVTSAPDTRAPAG--GSTAPRANG 946
834 STKPKTKMPKRVKKTPTTPPKMTSTMBELNP-----TSRIAEMIQOT--TPRP--NOTPN 885
947 VTSAPDTRP--APGSTAPRANGVTSAPDTRAPAGSTAPRANGVTSAPDTRAPAGSTAPR 1003
886 SKLVEVNPKSEDDAGABEPTPHMLLRHVMPETVTPMDKTPKRPVNGGIIINPMLSDET 945
```

Db 1004 AHGVTAPDTRPAPGSTAPPAAH-----CVTSAPDNRRALGSTA-----PVIHNVTS 1049

Qy 946 ICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPSPPS 981

Db 1050 ASSGASGASATLVHNGTNSARATTTTAPASKSPFISPS 1085

RESULT 9

T18535

high molecular mass nuclear antigen - chicken (fragment)

C:Species: Gallus gallus (chicken)

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #extl_change 15-Oct-1999

C:Accession: T18535

R:Shimada, K.; Harata, M.; Mizuno, S.

J. Cell Sci. 110, 3031-3041, 1997

A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick

A:Reference number: Z18955, M01D:9803440

A:Accession: T18535

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1151 <SH1>

A:Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BAA24137.1

Query Match 12.4%; Score 671; DB 2; Length 1151;

Best Local Similarity 25.9%; Pred. No. 3.9e-22;

Matches 284; Conservative 100; Mismatches 419; Indels 292; Gaps 55;

Qy 14 TPKP-----PVYDEAGSLDNGDFKVTTPDSTQHNVSTG--PKITTAKPINDRPS 64

Db 113 TPPESSQGPAGPPPSQGAAGAPKGDGTAPSGTSGADGKPAADVDVKATTA--ATEARP- 170

Qy 65 LPPNSDTSKETSITLVNKETTYETKETTNNK--QTSIDGKEKTSIAKETGSIETSKDL 122

Db 171 -----ASAASPTVPKATAEATAVTAASQAPKAADAAAVTA--SOSAPKATV-EV 219

Qy 123 APTSKVLAKPPKAEKTTTGGFALTPKPEPP--TPPKPEASTPKPE--TPPTIKSAPT 177

Db 220 KPAANAAYAKKAVTAAAKKATAEKAPAVTSPTIPCSSAEKPLTAASPTSKA--T 277

Qy 178 TPKPEAPT-----TKSAPTPPKPEAPTTPK-----203

Db 278 AEAKPVPATASLMTAKVTAEKAPSPSPKATTDTKAVTATADKAGPDVKAFAVCAEA 337

Qy 204 EPAPTTPKEPAPTTPKPEAPTTPKSAPT-----PKPEA-----PTTPPKKAPPTPK 250

Db 338 KPAPPPPPQOOLPKAAAAAAPTGTGLKPRATAPHGSPRANSHVTVTTPVNVKRAAAATVP- 396

Qy 251 EPAPTTPKEPPTT-----PKPEAPTTPKEPAPTTPKPAKPAKPAKPAKPAKPAK 306

Db 397 -TAGAVPRASGTTPPAAPQOPVP---KAAPVTPPSPOQAVPRAATAA---APVTPQO 448

Qy 307 P---APTTPKESPTTPKPEAPTTPKSAPTTKKAPAP---TTTSAPTTPKESPTTPKE 360

Db 449 PVTKAATTTNATPPQOPLPKAATTTTATVPVPOQPIPKAGDAAAPPVAPAPSDGGAAT 508

Qy 361 P---APTTPKPEAPTTPKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 406

Db 509 PGVPMATADPOKPPPTPOSVSATTEPKQORAPAPSPENKATPAVSPSPKLKSLPTIP 568

Qy 407 K-----EPAPTTPK---ETAPTTPKLPPTPEKLAAPTPPEKAPPTPELA----- 450

Db 569 KPVPLMALTPQPVTAQMTQLAATKPSPIVFKASER-ALMTPPPPPPPGLPRLAALAKLLG 627

Qy 451 -PTTP-----EETPTTP-----EPAPTTP-----KKAAPNTPK----- 480

Db 628 LPSSSVASAMNAKVTPLPLASVPVMAASPSLGDAAKVALAINAASPGAKPELAGNG 687

Qy 481 ----PAPTTPKEPAPTTPKPEAPTTPKETAAPL-----TPKGAAPT----- 516

Db 668 TLMAPMGAANTOMARIGAAGAAQTAPMGAANTHVSPMGAGATQMSPTGAANTHMSPIGA 747

Qy 517 -----TLKEPAPTTPKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 555

Db 748 GGATQMSPMGANTQMSPMGATTTQMSPMGAATQPSPMGAATQVATISAGNTQVSP 807

Qy 556 KG--TAPTTPKPEAPTTPKPEAPTTPKGAAPTTPKPE--APTTPKPAKPAKPAKPAK 611

Db 808 MGAATPPQTPSVGAATTP-QPSPM-----GAAATLMSPMGAATTPQ---PSPGAVTTP 859

Qy 612 ----TSTSDKP-APTTPKET--APTTPKEP--APTTPKPKP-APTTPPPPP----- 652

Db 860 PMAATNTTTPPPMAASTPOSTPMGAATTTQSPPMKATTTQSPPMKASPPQAPVAGSPT 919

Qy 653 PTTSEVSTPTTPKEPTTIHKSDESTPEL--SAEPTPALENSPEKGVPTTKTPAAKRP 710

Db 920 PEPPIRPSPTAQTSQPMKSPPPPPKAPSAAOISPAHAHVAASPGV-TAVSPA---P 975

Qy 711 EMTTAKDKDTEERDLRTPEPTTAAPKMI-KETATTTETTESKTLATTTQVSTTTQDT 769

Db 976 IGVTASPSADGARLSPGFPAATDGRKASPAATADVTETAAFD--VTAATAVPA----- 1029

Qy 770 TPKKITTLKTTTLAPKVTTTKKTTTTEIEMNKPEETAKPRDRAATNSKATTPKPKPTAP 829

Db 1030 AP-----TKAKRSS 1077

Qy 830 KKPTSTKKPKTMRVRKPKTTPPKRMISTMBELNFTSIAEAMLOTTTRPNOTNSKLV 889

Db 1078 -----PAVGDCQOQMTPGAOSVPP-----VTBAAYO----- 1104

Qy 890 EVNPKSEADGAGCE 904

Db 1105 EAAAAAAGAGERE 1119

RESULT 10

T25697

hypothetical protein F16F9.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #extl_change 18-Feb-2000

C:Accession: T25697

R:Fullon, B.

submitted to the EMBL Data Library, August 1996

A:Description: The sequence of C. elegans cosmid F16F9.

A:Reference number: Z20071

A:Accession: T25697

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1229 <FOL>

A:Cross-references: EMBL:U67956; PIDN:AA07691.1; GSPDB:GN00028; CESP:F16F9.2

A:Genetics:

A:Gene: CESP:F16F9.2

A:Map position: X

A:Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 12.2%; Score 659.5; DB 2; Length 1229;

Best Local Similarity 29.0%; Pred. No. 1.3e-21;

Matches 307; Conservative 78; Mismatches 339; Indels 333; Gaps 52;

Qy 33 KVTTP-PDTSITQ---HNKVSFSPKITTVA-KPINFRLSPNSDT-----SKETSITVN 80

Db 119 KVTTSASTNATNATTKGDSSTPEIITGIIVINSKSESVDMSSTFRSTLSPTTEILTS 178

Qy 81 KETTVETKETTNNKQTSSTGKEKETSIAKETOSIETKSAADLAPTSKVLAKPPKAEETT 140

Db 179 PETLVSTDSSTST-EQTSPPNTTEIASPMETNTTTEATTTTSEVSSTLAS---EDETTV 234

Qy 141 KGPAITTPKEPTTPTPKPEASTPKPEPTPTTIKSAPTTPKPEAPTTPKSAPTTPKPEAPT 200

Db 235 TAAEST-----TTVAEVSITTEETPTT-----AESTKKSTT-----KAPA 272

Qy 201 TTKEPAPTTPKPEAPTTPKPEAPTTPKSAPTTPKPEAPTTP-----KKPAPTTPKPEAPT 256

Db 273 TTPEPTPTTTEE--VTTEAETSTTTSSETSTEK---PTPLIDNKTAGATGK---PET 324

A:Map position: 2q31-2q31

Query Match	11.7%;	Score 632.5;	DB 2;	Length 7962;
Best Local Similarity	22.9%;	Pred. No. 1.1e-19;		
Matches 295;	Conservative 105;	Mismatches 453;	Indels 437;	Gaps 60;

[illegible][illegible]

```

RESULT 13
C84672
hypothetical protein At2g27380 [imported] - Arabidopsis thaliana
C.Species: Arabidopsis thaliana (mouse-ear cress)
C.Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C.Accession: C84672
R.Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bonito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Mofeet, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, L.; Tallon,
Guss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-766, 1999
A.Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A.Reference number: A84420; MUID:20083487
A.Accession: C84672
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-761 <STO>
A.Cross-references: GI:AE002093; NID:g5306260; PIDN:AA041992.1; GSPDB:GM00139
C.Genetics:
A:Gene: At2g27380
A:Map position: 2

```

Query Match	11.7%	Score 632:	DB 2:	length 761:
Best Local Similarity	31.1%	Pred. No. 1.3e-20:		
Matches 211; Conservative	44;	Mismatches 354;	Indels 82;	Gaps 33;

[illegible]


```

RESULT 14
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favellio, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Intervals: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2; 3504/1

```

Query Match	11.6%	Score 629;	DB 2;	Length 3507;
Best Local Similarity	24.1%	Pred. No. 7.2e-20;		
Matches 252;	Conservative 158;	Mismatches 416;	Indels 218;	Gaps 45;
QY	10	KKPKPKPPVNDVDEAGSDN-GDFKVTY-----PDNSTGHNNVSNRSPKLTITA-----K	57	
Db	1942	KKPKPKREIIDEEMSSSSNSQGEKPTTKGIVSSTSATSSESTTAEPHTTSSISSTTSK	2001	
QY	58	PINRPSLPPN---SDTSKETSLEYNKETTV-ETKETTNNQTSYDGEKKTSAKETOS	113	
Db	2002	DMTSSKS-PEWMTSESSEPEVSTSSSKSTTASSETTVSSRPSSSSESEALUTSSPATTEV	2060	
QY	114	IEKTSAKLAP---TSKVLAKPTPKAETTKGPAALTTEKPTPTPKBPASTTKEPP	169	
Db	2061	ITESSKVTTPKEBSSSEITVTKLSSPEVTESSVKSSSTPS-TTQSVSTSVTPETSKS	2119	
QY	170	TTIKS-APTTPKEPAPTTKSAPTPPKBPAPTTTKEBAPTTTKEBAPTTTKS	228	
Db	2120	TVLSSEAPVTSSTPEVHT-SSETKPSSLASSTGGTNNSTSTSSLASVKTSADGTS	2178	
QY	229	APTTPKEBAPTTPKKBPATTPK-BEAPATTPKEBPTTTPKEBAPTTKEP-APTTPKEBAP	286	
Db	2179	ASVAPVKLSSLPDVGSGSTKTFDATTESSTVQASETSOGTSVKSTSEPSHNTKLSTSS	2238	
QY	287	APKBPATTPKEBAPTTTKEBAPTTTKEBSPPTP--KEBAPTTTKSAPTTTKEBAPTTTK	344	

```

Db      2239  NQSSSVPTVSPKSTVPPE-----STQEPSTTTPSQQLTTPMKNSEVLTTSEPHLSSS  2293
OY      345    SAP-----TTKKPEPPTTKEBPATTTKEBPATTTKKAPATTTKEBPATTTKEBPATTT  397
Db      2294  LSPDVQSSTTNNLSESTVE---TPKTSSEVSLNSEBPSTTEAPTLSPDILSTTN  2349
OY      398    -----TKKPATAPK-----EBAPTTKETAAPTTK  427
Db      2350  NLQSSTVEDRSELSENSEKPT-SAPELVTSSTVHVASSPDVPTES--SEPDLG  2406
OY      428  TTPEKLAPTTPEKPAPTTPEELAPTTPEEPPTTPEEP-----APTPKAAAPN  476
Db      2407  SSTENPEASSKQTISSSTPTPDTTTASEBPTKSTSMSPDLSTTNVLSSESTPRESS--  2464
OY      477  TPKEBAPT-----PKBPAPT-----TPKEBAPT  510
Db      2465  --KSPVSSKEGIVVSTSEKVPESITISVLEEDLKTTPSPILEETTTASESEPLT  2522
OY      511  KGTAPTTKEBPATTTPKKPAKELAPTTTKEPTSTSDKA-----PT-----  553
Db      2523  EDSLAVYRIHELTTSENVKPESESTTSSSESKNSQBPALITSTVYVPISSVLTA  2582
OY      554  -----TP--KGTAP--TTPK-----EBAPTTKEBPATTTKGTAPTTKEBAPT  593
Db      2583  SEIEATISNTPKQORTPTTITSPKSLVKTSPSTVTSSEPESESTKRTTSTVSTTPT  2642
OY      594  TPKKAPKELAPTTTKGPP-STTSOKPAPTTPKEKETAAPTTPKBPATTTPKKAPTTPEPP  652
Db      2643  EBTTSSESLTLAASKPRESTTESSEAPTTTAKTSSEKPSNVASTSRKSTENVETSTQ  2702
OY      653  PTTSEVSTPTTKEPTTIHKSPDESTR--ELSAEPTPALE--NSPKKEGVPTTTPPAT  708
Db      2703  SGLSESTMSSTSS-----SEBETNAPAVTSSSEASTLLEENSTSSP---TSSASV  2752
OY      709  K-----PEMTTAKDKTERDLRTPPETTTAAPKMTKETATTTETTESKITATTT-----  759
Db      2753  KLSLPEPESITSEAVTSSR---APAEITMSSEKREISTVSSSEPEEIPPLSTVSPN  2808
OY      760  -----QVSTTTPQDTPP-FKITPP-----LKTTLAPKVTTTKTITTEIMKPP  802
Db      2809  VYTASSIPSEEPILSSVSSSTPRVRLTTCPPDDLIVSYVPSHKNRQNTIASSV---P  2865
OY      803  EETAKPKORATNSKATTPPKPOKPTKAPKKPTSTKPKTMAPRYRKPK---TTPTPKMTST  859
Db      2866  SNSSTPILPSES-LTTPQPPPTTTTAKKPAATTSKRGPPSIQBPAAEMFTTPAP-----  2918
OY      860  MPELN-----PTSRIAEAMLOTTT  878
Db      2919  PPSNNGYGEETNOEEOVSTTT  2942

```

RESULT 15

I51618

nucleolar phosphoprotein - African clawed frog

C:Species: *Xenopus laevis* (African clawed frog)

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999

C:Accession: I51618; S57757

R:Caillars, C.; McStray, B.

J:Cell_Sci. 108, 3339-3347, 1995

A:Title: Identification and cDNA cloning of a *Xenopus* nucleolar phosphoprotein, xNopp

A:Reference number: I51618; M01D:96019267

A:Accession: I51618

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-990 <CAI>

A:Cross-references: EMBL:X08927; NID:g895920; PIDN:CA61368.1; PID:g895921

C:Genetics:

A:Gene: xNopp180

C:Superfamily: nucleolar-cytoplasm shuttle phosphoprotein

C:Keywords: phosphoprotein

Mon Apr 29 08:35:37 2002

us-09-556-246-1_copy_200_1212.rpr

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:19:50 ; Search time 47.4 seconds

(without alignments)
783.575 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1212
Perfect score: 5416
Sequence: 1 VKDNKKRRTKKRTPKPPVY.....SPIDVFTRCNCEGKTFEKK 1013

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 100059 seqs, 3664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	1168	21.6	5179	1	MUC2_HUMAN
2	942	17.4	1664	1	SLPI_CLOM
3	789.5	14.6	1367	1	AMYH_YEAST
4	651	12.0	1255	1	MUC1_HUMAN
5	556.5	10.3	875	1	FPI_MYTE
6	555.5	10.3	2700	1	ZAN_HUMAN
7	551	10.2	620	1	EXTN_TOBAC
8	533	9.8	1087	1	NFH_MOUSE
9	530.5	9.8	1162	1	TCNA_TRYCR
10	530	9.8	865	1	CPN_DROME
11	518.5	9.6	872	1	FPI_MYTCO
12	509.5	9.4	662	1	MUC1_XENLA
13	503.5	9.3	1970	1	RPB1_HUMAN
14	498.5	9.2	1970	1	RPB1_MOUSE
15	497	9.2	831	1	NFH_RAT
16	493.5	9.1	467	1	RPB1_CRIGR
17	488.5	9.0	826	1	SSP2_PLAYO
18	488	9.0	1020	1	NFH_HUMAN
19	475.5	8.8	267	1	EXTN_MAIZE
20	471.5	8.7	5376	1	ZAN_MOUSE
21	468.5	8.7	634	1	HMP1_CANAL
22	467	8.6	817	1	VREP1_YEAST
23	463	8.5	2142	1	BAT2_HUMAN
24	454	8.4	797	1	VGLX_HSEB
25	454	8.4	1161	1	V39P_YEAST
26	448.5	8.3	670	1	VG50_HSV1
27	442.5	8.2	751	1	FPI_MYTCO
28	439.5	8.1	1083	1	T2D3_HUMAN
29	432.5	8.0	1439	1	XP2_XENLA
30	432	8.0	3164	1	TEGU_HSV1
31	426	7.9	1275	1	TRX2_HUMAN
32	424.5	7.8	1125	1	MAP4_MOUSE
33	424	7.8	307	1	SGS3_DROME

34	424	7.8	1251	1	YOU3_CAEEL
35	422.5	7.8	1794	1	VAV1_SCHPO
36	422.5	7.8	3421	1	TEGU_HSV1B
37	419.5	7.7	2476	1	ZAN_PIG
38	419.5	7.7	2774	1	MAPA_RAT
39	419	7.7	1185	1	DRPL_HUMAN
40	415	7.7	1229	1	N121_HUMAN
41	411.5	7.6	907	1	VGP3_EBV
42	410	7.6	1183	1	DRPL_RAT
43	407.5	7.5	3256	1	K167_HUMAN
44	405.5	7.5	2517	1	NCR2_HUMAN
45	405	7.5	1411	1	TCOF_HUMAN

ALIGNMENTS

RESULT	ID	MUC2_HUMAN	STANDARD:	PRT:	5179 AA.
AC	002817	014878;			
DT	01-JUN-1994	(Rel. 29, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DT	20-AUG-2001	(Rel. 40, Last annotation update)			
DE	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).				
GN	MUC2 OR SMUC.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
OX	NCBI_TaxID=9606;				
RP	SEQUENCE FROM N.A.				
RP	TISSUE=Intestine;				
RC	MEDLINE=94132002; PubMed=8300571;				
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;				
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.				
RT	Identification of the amino terminus and overall sequence similarity				
RT	to prepro-von Willebrand factor.";				
RL	J. Biol. Chem. 269:2440-2446(1994).				
RN	[2]				
RP	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.				
RP	TISSUE=Colon;				
RC	MEDLINE=93016075; PubMed=1400449;				
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,				
RA	Kim Y.S.;				
RT	"The human MUC2 intestinal mucin has cysteine-rich subdomains located				
RT	both upstream and downstream of its central repetitive region.";				
RL	J. Biol. Chem. 267:21375-21383(1992).				
RN	[3]				
RP	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.				
RC	MEDLINE=91358717; PubMed=1885763;				
RA	Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,				
RA	Petersen G.M., Kim Y.S.;				
RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays				
RT	and polymorphism.";				
RL	J. Clin. Invest. 88:1005-1013(1991).				
CC	-1- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND				
CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A				
CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIONS				
CC	AGENTS AT MUCOSAL SURFACES.				
CC	-1- SUBUNIT: MULTIMERIC.				
CC	-1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,				
CC	BRONCHUS, CERVIX AND GALL BLADDER.				
CC	-1- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR				
CC	INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).				
CC	-1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND				
CC	VARIES AMONG DIFFERENT ALLELES.				
CC	-1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT				
CC	OF SILKORM HEMOCYTIN.				
CC	-1- SIMILARITY: CONTAINS 2 WFCC DOMAINS.				
CC	-1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).				

CC This SWISSPROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).
CC or send an email to license@sib-sib.ch).

CC	-	-	-
DR	EMBL; L2198;	AAB95295.1;	-
DR	EMBL; M74027;	AAA59875.1;	-
DR	EMBL; M94131;	AAA59163.1;	-
DR	EMBL; M94132;	AAA59164.1;	-
DR	MIM; 158370;	-	-
DR	InterPro; IPR000359;	Cys_knot.	
DR	InterPro; IPR000561;	EGF-like.	
DR	InterPro; IPR002400;	GF_cysknot.	
DR	InterPro; IPR001007;	WMFC.	
DR	Pfam; PF00007;	Cys_knot; 1.	
DR	Pfam; PF00094;	wvd; 4.	
DR	PRINTS; PR00438;	GFCYSKNOT.	
DR	SMART; SMO0214;	VMC; 2.	
DR	SMART; SMO0011;	VMC_def; 2.	
DR	SMART; SMO0216;	WVD; 4.	
DR	PROSITE; PS00022;	EGF_1; UNKNOWN_1.	
DR	PROSITE; PS01185;	CTCK_1; 1.	
DR	PROSITE; PS01225;	CTCK_2; 1.	
DR	PROSITE; PS01208;	WMFC; 2.	
KW	Glycoprotein;	Repeat; signal.	
FT	SIGNAL	1	POTENTIAL.
FT	CHAIN	21	MUCIN 2.
FT	DOMAIN	1401	APPROXIMATE REPEATS.
FT	REPEAT	1401	1.
FT	REPEAT	1417	1.
FT	REPEAT	1417	2.
FT	REPEAT	1433	3.
FT	REPEAT	1433	3.
FT	REPEAT	1448	2.
FT	REPEAT	1449	4.
FT	REPEAT	1465	4.
FT	REPEAT	1465	5.
FT	REPEAT	1471	5.
FT	REPEAT	1472	6.
FT	REPEAT	1478	6.
FT	REPEAT	1479	7A.
FT	REPEAT	1494	7B.
FT	REPEAT	1495	7B.
FT	REPEAT	1517	8A.
FT	REPEAT	1518	8B.
FT	REPEAT	1534	8B.
FT	REPEAT	1536	8B.
FT	REPEAT	1557	9A.
FT	REPEAT	1572	9A.
FT	REPEAT	1573	9B.
FT	REPEAT	1597	10A.
FT	REPEAT	1612	10A.
FT	REPEAT	1613	10B.
FT	REPEAT	1635	10B.
FT	REPEAT	1636	11A.
FT	REPEAT	1651	11A.
FT	REPEAT	1652	11B.
FT	REPEAT	1676	12.
FT	REPEAT	1683	12.
FT	REPEAT	1684	13.
FT	REPEAT	1699	13.
FT	REPEAT	1700	14.
FT	REPEAT	1715	14.
FT	REPEAT	1716	15.
FT	REPEAT	1731	15.
FT	REPEAT	1732	16.
FT	REPEAT	1747	16.
FT	DOMAIN	4815	WMFC 1.
FT	DOMAIN	4924	WMFC 2.
FT	DOMAIN	4991	CTCK.
FT	DOMAIN	5075	BY SIMILARITY.
FT	DISULFID	5075	BY SIMILARITY.
FT	DISULFID	5089	BY SIMILARITY.
FT	DISULFID	5098	BY SIMILARITY.
FT	DISULFID	5152	BY SIMILARITY.
FT	DISULFID	5102	BY SIMILARITY.
FT	DISULFID	? 5159	BY SIMILARITY.
FT	CARBOHYD	163	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	423	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	423	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	670	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	670	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	770	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	894	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1139	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1139	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1154	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1215	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1230	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1230	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1246	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1246	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1787	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1787	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1820	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1820	N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	4339	4339	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	4351	4351	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	4362	4362	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	4373	4373	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	4422	4422	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	4438	4438	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	4502	4502	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	4616	4616	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	4627	4627	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	4752	4752	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	4787	4787	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	4881	4881	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	4888	4888	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	4955	4955	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	4970	4970	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	5019	5019	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	5038	5038	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CARBOHYD	5069	5069	N-LINKED	(GLCNAC. .)	(POTENTIAL.)
FT	CONFLICT	1351	1351	H -> L (IN REF. 3).		
FT	CONFLICT	1412	1412	T -> S (IN REF. 3).		
FT	CONFLICT	1449	1449	L -> P (IN REF. 3).		
FT	CONFLICT	1504	1504	M -> T (IN REF. 3).		
FT	CONFLICT	4192	4192	G -> S (IN REF. 2).		
SO	SEQUENCE	5179	AA: 340295	MM: 85CD5751FB9A5663	CR6C4:	

Query Match	21.6%;	Score 1168;	DB 1;	Length 5179;
Best Local Similarity	32.5%;	Pred. No. 1.4e-39;		
Matches 363;	Conservative 53;	Mismatches 448;	Indels 254;	Gaps 40;

```

0Y 21 DEAGSGLDNGDPK-----YTPPDTSTTQH-NKYSTSPK-----52
Db 1312 DHPSSGSDDGREPPDVGCGAGEDIECRSVXDPHLSLBHOQKQVODVSVGFIKNEQDF 1371
0Y 53 -----ITPAKKIPNRPRLSPNDSPTSKESLITVNEKETVETK 88
Db 1372 GNGPGLCYDYKIRNCCWPMKDLTTPSSPPTTSSPPTTTTLPLPTTSSPPT-----1427
0Y 89 ETTTNNKOTSTDGKERKTTSAKETOSIEKTSAKDLAPTSKVLAKPPKAEITTKGALITP 148
Db 1428 -TTTTPPPTTSPPTSPPTTTTTTP-----LPTT---TPSPPISTTTTPPTTTP 1470
0Y 149 KEPT-----PTPKBPASATPEKEPTTIIKSAF-TTPKEP-----APTTPKSAF-TTP 194
Db 1471 SPPTTTPSPPTTSSPPTTTTTTTPPTTTPSPSPMTIPLTPASTTLLEPLTTPSPPTTTT 1530
0Y 195 KEPAPTTEKAPATTPKEP-----APTTEKAPATTTKSAF-TTPKEAPITTPKAPA 245
Db 1531 TTPPTTTPSPPTTTPITTPSTTLPLPTTTPSPPTTTPPTTTPPTTTPPTTTPSPPT 1590
0Y 246 PTTKEBAPTPEKEPTPTTPKEBAPTTEBAPTTPKEP-----APAPAKKAPATTPKE 298
Db 1591 TTTTTPPTTTPSSPPTTTTTTTPPTTTPSPPTTTTITPPTSTTLPLPTTTPSPPTTTT 1650
0Y 299 PAPPTPKBAPTPEKAPTPEBPTTPKEBAPTTPKSAPTTEKAPATTTKSAF--TTPKEBSP 355
Db 1651 PPTTTPSPPTTTPSPPTTTTTTTPPTTPPTTSSPITTPSPPTTMTTPSPPTTSSPPT 1709
0Y 356 TTTKEBAPTPEKEBAPTTPKKAPITTPKEBAPTTPKEBAPTTPKKAPATAKREBAPTTP- 414
Db 1710 TTTTTPSPSTTTPSPPTTMTTPSPPTTSSPPTTTPMTTLPLPTTSSPLTTTTLPLPTPT 1769
0Y 415 -----KETAPTTP-----KKLETP 427
Db 1770 ESPSTTTPPTTCVELCNMTGMLDSGKRNFKRGSDTELIGVCGPGMAANISCRATMYP 1829
0Y 428 -----TTPKEKLAPTT 437
Db 1830 DVPIGOLGOTVACDVSVGLICKNEQDQGVIPMAFCLNYPEINVOCCQCVQOPTMTMTT 1889
0Y 438 PKKAPATPEELAPATTPPEBPTT-TPBEPAPTTPKAAPNTPKEBAPTTPKEP-----AP 491
Db 1890 TNPAPTPTTTLTTTTPVTPPTPTGOTPTTTLITTTTTVTPPTPTGOTPTTTLPTT 1949

```

QY	492	TPKRPAPL--PKKEAPLTPPKSTAPLTLKEAPLTPPKKAPKAPKELAPLTPPKKSTSDSKP	550
Db	1950	TTTTVTPRTPTGQTPRTTPTTPTTPTTTPVTPRPTPTGQTPPTTPTTPTTTPVTPRPTPTGQ	2009
QY	551	AP-TTPKSTAPRTTPKEAPATTPKEBAPRTTPKSTAPRTTLKEAPATTPKKEAPKELAPTTTK	609
Db	2010	TPRTTPTTTTTVTTPRTPTPTGQTPRT--TPRTTTTTVTPRPTPTGQTPPTTPTTTTTV	2068
QY	610	GPTSTTSOKRAP--TPKKEAPLTPKKEBAPRTTPKKAAPTTPETPRPTTSVSPTTTKEPT	668
Db	2069	TPRTPTPTGQTPRTTPTTPTTTPVTPRPTPTGQTPRT--TPRTTTTTVTPRPTPTGQTPRT	2127
QY	669	TIHKSPDSTPRLSAPRTPKALENSPKPE-----GVPR--TKPKAPTPKEMTTAK	717
Db	2128	T---TPRTTTTTVTPRPTPTGQTPRTTPTTPTTTPVTPRPTPTGQTPPTTPTTTTTVT	2184
QY	718	DKTERDLR--TPP--ETTPAPKMT--KEPATTEKTESKITATTTQVSTTTQDT	769
Db	2185	PTPTPTGQTPPTTPTTPTTPTTTPVTPRPTPTGQTPRTTPTTPTTTTTVTPRPTP--TGQTPPT	2243
QY	770	TPFKTTTLKTTTLAKVTT--KKTITTEIIMKKPEETAKPKRAINSKATTPKPKQPKTA	828
Db	2244	TP--TTT--TTTTPPTPTPTGQTPPTTPTTPTTPTTPTTTPVTPRPTPTGQTPPTTPTTTTTVT	2299
QY	829	P-KRTSTKPKPTMTRVKKRPTTPPKMTSMPELN---TSRLAEMALQTTTR--PNQT	883
Db	2300	PTPTPTGQTPPTTPTTPTTPTTTPVTPRPTPTGQTPRTTPTTPTTTTTVTPRPTPTGQTPPT	2359
QY	884	PNSKLVEVNPKSEDAGAEGETPHMLRPHVMEVTP	921
Db	2360	PTTTTTVTPRPTPTG---TQTP--TTTPRTTTTTVT	2392
RESULT	2		
SLP1	SLP1_CLOTM	STANDARD:	PRT: 1664 AA.
AC	006852:		
DT	01-JUN-1994 (Rel. 29, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	01-FEB-1996 (Rel. 33, Last annotation update)		
DE	CELL SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER PROTEIN 1).		
GN	OLBP.		
OS	Clostridium thermocellum.		
CC	Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.		
CC	NCBI_TaxID=1515;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NCIB 10682;		
RX	MEDLINE=93209931; PubMed=8458832;		
RA	Fujino T., Begun P., Aubert J.-P.;		
RT	"Organization of a Clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein CfpA and a protein possibly involved in attachment of the cellulosome to the cell surface.";		
RL	J. Bacteriol. 175:1891-1899 (1993).		
CC	-1. SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.		
CC	-1. SUBCELLULAR LOCATION: CELL WALL.		
CC	-1. SIMILARITY: CONTAINS 4 S-LAYER HOMOLOGY (SLH) DOMAINS.		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/).		
CC	or send an email to license@isb-sib.ch .		
CC	-----		
DR	EMBL: X67506; CAA47841.1; --		
DR	InterPro: IPR001119; SLH.		
DR	Pfam: PF00395; SLH; 3.		
DR	PROSITE: PS01072; SLH_DOMAIN; 2.		

[illegible]

DB	1425	-----	KPSTSTAPPEIE-EPTPSDYPGALGGEHAYLNGY-----	PDGSRP	1465
oy	871	EMBL	QTTFRPQGNFSKLYEVNRPSEADSGEGENPHMLLRPHVEMPEVTEMDMPLP	927	
Db	1425	-----	KPSTSTAPPEIE-EPTPSDYPGALGGEHAYLNGY-----	PDGSRP	1465
RESULT	3				
AMYL_YEAST		STANDARD;	PRT: 1367 AA.		
ID	AMYL_YEAST				
AC	P08640; P08068;				
DC	01-AUG-1988 (Rel. 08, Created)				
DT	01-FEB-1995 (Rel. 31, Last sequence update)				
DR	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	GLUCOAMYLASE SI/LS2 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-				
DE	GLUCOSIDASE)				
GN	STAL OR STA2 OR MA15 OR YTR019C.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
OX	NCBI_Taxid:4932;				
PN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-S288C / AB972;				
RA	Barrell B.G., Badcock K., Bankier A.T., Bowmen S., Brown D.,				
RA	Chargier C.M., Connor R., Copesey T., Dear S., Devlin K., Fraser A.,				
RA	Geisler S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,				
RA	Leites E., Lye G., Moule S., Moule T., Odell C., Pearson D.,				
RA	Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,				
RA	Walsh S.V., Whitehead S.;				
RL	Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.				
RN	[2]				
RP	SEQUENCE OF 1-242 AND 762-1331 FROM N.A.				
RX	Medline=87194600; PubMed=3106330;				
RA	Yamashita I., Nakamura M., Fukui S.;				
RT	"Gene fusion is a possible mechanism underlying the evolution of				
RT	STAL.";				
RL	J. Bacteriol. 169:2142-2149(1987).				
RN	[3]				
RP	SEQUENCE OF 1-31 FROM N.A.				
RC	STRAIN-Spx101-1C.				
RA	Medline=89031230; PubMed=3141213;				
RT	Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;				
RT	"Similar short elements in the 5' regions of the STA2 and SGA genes				
RT	from Saccharomyces cerevisiae.";				
RL	FEBS Lett. 239:179-184(1988).				
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-				
CC	GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS				
CC	WITH RELEASE OF BETA-D-GLUCOSE.				
CC	-1- SIMILARITY: TO S.POMBE SPPC215.13.				
CC	-1- SIMILARITY: SOME, TO S.POMBE SPPC285.13C.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL, Z38061; CAA86176.1; -				
DR	EMBL, M16164; AAA35014.1; -				
DR	EMBL, M16165; AAA35015.1; -				
DR	EMBL, X13857; CAA32069.1; -				
DR	PIR, B26877; B26877.				
DR	PIR, A26877; A26877.				
DR	PIR, S48478; S48478.				
DR	SGD; S0001458; M0C1.				
KW	Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;				
KW	Signal; Multigene family.				
FT	SIGNAL	1	21	POTENTIAL.	
FT	CHAIN	22	1367	GLUCOAMYLASE SI/S2.	
FT	DOMAIN	210	1367	SER/THR-RICH.	
FT	CARBOHYD	817	817	N-LINKED (GLCNAC. . .) (POTENTIAL).	

FT	CARBOHYD	874	874	N-LINKED (GLCNAc...)	(POTENTIAL)
SO	SEQUENCE	1367	AA: 136110	MM: 91C00E2DBD61AA9D	CRC64:
	Query Match	14.6%	Score 789.5;	DB: 1;	Length 1367;
	Best Local Similarity	28.4%	Pred. No. 4.6e-25;		
	Matches 301;	Conservative 104;	Mismatches 483;	Indels 171;	Gaps 47;
QY	35	TTPTDSTQHNKAVSPKITTAKPLNP-----RPSLPRNSDPKSTSLVNKETEVE	86		
DB	236	STSESTTSSSTSSSTSSSTTAPPTPTTSCYKKKPLPPTTSCYKPKPPPHHDTPPC	295		
QY	87	TKETITTKQSTDOKEKETSIAKETOSIEKTSAKNDLAPTSKVLAKVPKAEYTT--TKCPA	144		
DB	296	TKKTKTTSK-TCT---KKTTTPVPPTPS--STTSSSSAPV-----PTSSSTTSSSSAPV	344		
QY	145	LITRE-----PPTTKEPASTPKKEPTPTTIKSAF-----TKPEKAPITTKAPAPT	194		
DB	345	TSTSTSSSAPVPPTSSSTSSSSAPVTSSTSSSSAPVTSSTSSSSAPVPPTSSSTTE	404		
QY	195	KEAPATTTKEAPATTPKKEAPATTTKEAPATTTKAPITPKKEAPATTPPKAPATTKKEAP	254		
DB	405	SSSAPVTS-----TSSSSAPVTS--STSSSSAPV-----SSSTSSSAPVTSSTTES	453		
QY	255	TTPKKEPT---TTPEKAPITTKKEAPATTPKKEAPATPKKAPATTKKEAP-----TTPEK	307		
DB	454	SSAPVPPTSSSTTSSSSAPV-----STSSSSAPV--VPPTSSSTTSSSSAPVTSSTSSS	509		
QY	308	APTTKEPSPTPKPEAPATTTKSAPTTTKEAP---TTTKAPITPKKEPSPTTKKEAP	363		
DB	510	APVPTPS--STTSSSSAPAPPTSSSTTSSSSAPVTSSTSSSSAPVPPTSSSTTSSSNP	568		
QY	364	---TTPEKAPAT--TPKKAPATTKKEAPATTPKKEAPITTKKAPATLAKKEAPITPKETA	418		
DB	569	VTSSTSSSSAPVPPTSSSTTSSSSAPVPT---PSSSTSSSSAPAPPTSSSTSSSA	624		
QY	419	PTTKKLPTTPEKAPATTPKKEAPATTPPEELAPT--TPPEPTTPEEAP-----TPKA	472		
DB	625	PVT---STTSSSSAP--VPPTSSSTTSSSSAPVPTSSSTTSSSSAPVPTSSSTTSSS	679		
QY	473	AAPNTPKPEAPATTPKPEAPATTPKKEAPATTPKETAAT--TPKGTAPTTLKEAPATTPKKAP	531		
DB	680	SAPV-----STTSSSSAPV-----STTSSSSAPVPPTSSSTTSSSSAPVPPTSSSTTE	731		
QY	532	KELATTTKEPTSTSDKAPATTPKGTAPTTKKEAP---TPKKEAPAT--PKGAPATL	587		
DB	732	SSSAPVPPTPS--STTSSSSAPVTSSTTSSSSAPVPPTSSSTTSSSSAPVPPTSSSTTSSS	790		
QY	588	KEAPATTPKPKAPKALAPTTTGGPTSTSDKAPATTPKETAATPKKEAP---TPPK	642		
DB	791	SAPVPTPSSSTTSSSSAPVPPTPSSSNITSSAPSSPTPSSSTSSSSAPVPPTSSSTTSSS	850		
QY	643	PAPTTPEPTPTTSEVPT-----TKPEPTIHKSPDESTPELSAEPKALENSKEP	697		
DB	851	SAPVSSSTTSSSSAPVPPTSSSNITSSAPSS1---PSSSTTSSSTGT--IVTPSSSKP	906		
QY	698	GAPTTKTPAATPEKETTAKDKTTERDLATTPPETTTAPAKMKETATTEKTEKTIAT	757		
DB	907	GSQETSVSSTETIIVPKITTSVTPBSTIITTTVCGSTGNSAGETSGCSPVTVT	966		
QY	758	---TTQVSTSTTODTTPKIKITLKITTLAPKVT---TKKTIITT--EIMKKPEETAKPKDR	811		
DB	967	VPTTTTTSVTSSTIITTTVCGSTGNSAGETSGCSPKTIITTTVPCSTPETA-----	1021		
QY	812	ATNSAKATPKPKP-----TKAPKPTSTKPKPTMPVVRKPKATTPPR	854		
DB	1022	---SSSTTSTPTTPVTVVSTTVTTEISTSKPKGEIITITTVTKNIPITTLITLAPTP-	1077		
QY	855	KMTSTMPELNPSRIAEAMLQTTTPNQTNSKLIENVNPKSEDAGAEGETPHMLRPV	914		
DB	1078	-SVTIVTNTPTT--ITTVTCSGT-----NSAGETSGCSPKVTITTVTPCST	1122		
QY	915	EMPEVTPMDMYLPKVPNOCIIINPMALSDETNICNGKPVUDGLITLNGTLVATRGHYFWML	974		

DB 1123 GTGGTTEATTLLVTTAVTTTAVTTTSSSTGTNSA-GKTTTGTTKSVPT-----YVTTL 1175
 QY 975 SPFSPPSPARRITEWYGIPSDIFVTRC-----NCEGKT 1009
 DB 1176 APSAPVTPATN-----AVPTTITT--TECSAATNAGET 1207

RESULT 4
 MUC1_HUMAN STANDARD: PRT: 1255 AA.
 ID MUC1_HUMAN
 AC P15941: P15942, P13931, P17626; Q14128; Q16442; Q16437; Q914J2;
 DT 01-UN-1990 (Rel. 13, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM) (PEMT)
 DE (EPITHELIAL) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)
 DE (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H23AG) (PEANUT-
 DE REACTIVE URINARY MUCIN) (PUM) (BREAST CARCINOMA-ASSOCIATED ANTIGEN
 DE Df3).
 GN MUC1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eulheria; Plimates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=90368716; PubMed=2394722;
 RA Ian M.S., Batta S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
 RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
 RL J. Biol. Chem. 265:15294-15299(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=90202794; PubMed=2318825;
 RA Lightenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
 RT "Epistatin, a carcinoma-associated mucin, is generated by a
 RT polymorphic gene encoding splice variants with alternative amino
 RT termini.";
 RL J. Biol. Chem. 265:5573-5578(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=90368715; PubMed=1697589;
 RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
 RA Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
 RT "Molecular cloning and expression of human tumor-associated
 RT polymorphic epithelial mucin.";
 RL J. Biol. Chem. 265:15286-15293(1990).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=91097524; PubMed=2268309;
 RA Lancaster C.A., Peat N., Duhig T., Wilson D.,
 RA Taylor-Papadimitriou J., Gendler S.J.;
 RT "Structure and expression of the human polymorphic epithelial mucin
 RT gene: an expressed VNMR unit.";
 RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=90276413; PubMed=2351132;
 RA Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N.,
 RA Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,
 RA Keydar I.;
 RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
 RT may generate multiple protein forms.";
 RL Eur. J. Biochem. 189:463-473(1990).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=90276414; PubMed=2112460;
 RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
 RA Zrlhan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;

RT "A transcribed gene, containing a variable number of tandem repeats,
 RT codes for a human epithelial tumor antigen. cDNA cloning, expression
 RT of the transcribed gene and over-expression in breast cancer
 RT tissue.";
 RL Eur. J. Biochem. 189:475-486(1990).
 RN [7]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91033045; PubMed=1688329;
 RA Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,
 RA Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;
 RT "Isolation and characterization of an expressed hypervariable gene
 RT coding for a breast-cancer-associated antigen.";
 RL Gene 93:313-318(1990).
 RN [8]
 RP PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=88330762; PubMed=2417635;
 RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
 RA Burchell J.;
 RT "A highly immunogenic region of a human polymorphic epithelial mucin
 RT expressed by carcinomas is made up of tandem repeats.";
 RL J. Biol. Chem. 263:12820-12823(1988).
 RN [9]
 RP SEQUENCE OF 1-169 FROM N.A.
 RX MEDLINE=90088473; PubMed=2597151;
 RA Ade M., Siddiqui J., Kufe D.;
 RT "Sequence analysis of the 5' region of the human DF3 breast
 RT carcinoma-associated antigen gene.";
 RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
 RN [10]
 RP SEQUENCE OF 1-109 FROM N.A.
 RC TISSUE=Thyroid;
 RX MEDLINE=96183746; PubMed=8608966;
 RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;
 RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
 RT transcriptase polymerase chain reaction of the MUC1 gene.";
 RL Int. J. Cancer 66:55-59(1996).
 RN [11]
 RP SEQUENCE OF 1-89 FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=96181716; PubMed=8604237;
 RA Yu C.-J., Yang P.C., Shew J.-Y., Hong T.-M., Yang S.-C., Lee Y.-C.,
 RA Lee L.-N., Luh K.-T., Wu C.-W.;
 RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
 RT tissues.";
 RL Oncology 53:118-126(1996).
 RN [12]
 RP SEQUENCE OF 1-46 FROM N.A.
 RC TISSUE=Breast carcinoma;
 RA Bulwela L., Liu Q., Lugmani Y.A., Gonn J.J., Coombes R.C.;
 RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
 CC CYTOSKELETON.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM
 CC IS ALSO PRODUCED.
 CC -1- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE
 CC SPLICING.
 CC -1- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL
 CC TUMORS, SUCH AS BREAST CANCER.
 CC -1- PTM: HIGHLY GLYCOSYLATED (N-AND O-LINKED CARBOHYDRATES AND STALIC
 CC ACID).
 CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT
 CC VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE
 CC MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.
 CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: J05582: AAA60019.1; -
 DR EMBL: M32738: AAA35804.1; -
 DR EMBL: M32739: AAA35806.1; -
 DR EMBL: J05581: AAA59876.1; -
 DR EMBL: M61170: AAB53150.1; -
 DR EMBL: X52229: CA36477.1; ALT_SEQ.
 DR EMBL: X52228: CA36477.1; ALT_SEQ.
 DR EMBL: M35093: AAB59612.1; ALT_SEQ.
 DR EMBL: M31823: CA78972.1; -
 DR EMBL: M31825: CA78973.1; -
 DR EMBL: M31823: AAA55757.1; -
 DR EMBL: S81781: AAD14376.1; ALT_INIT.
 DR EMBL: S81736: AAD14369.1; ALT_INIT.
 DR EMBL: M21868: AAA39874.1; ALT_SEQ.
 DR PIR: A35175: A35175.
 DR PIR: B35175: B35175.
 DR PIR: S10218: S10218.
 DR GlycoSuiteDB: P15941; -
 DR MIM: 158340; -
 DR MIM: 113720; -
 DR InterPro: IPR000082: SEA.
 DR SMART: SMO0390: SEA; 1.
 DR SMART: SMO0200: SEA; 1.
 DR PROSITE: PSS0024: SEA; 1.
 KM Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
 Repeat: Alternative splicing.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 1 23 MUCIN 1.
 FT DOMAIN 24 1162 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1163 1186 POTENTIAL.
 FT DOMAIN 1187 1255 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 81 960 44 X 20 AA TANDDEM REPEATS.
 FT DOMAIN 1034 1151 SEA.
 FT CARBOHYD 957 957 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 975 975 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1029 1029 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1055 1055 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1133 1133 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VASPLIC 19 19 T-> TATTAKPRP (IN ISOFORM B).
 FT VASPLIC 20 22 MISSING (IN ISOFORM C).
 FT VASPLIC 20 31 MISSING (IN ISOFORM D).
 FT VASPLIC 126 905 MISSING (IN POLYMORPHIC EPITHELIAL ISOFORM).
 FT VASPLIC 1077 1087 FLOIRKGGFL -> VSLGSLPMLP (IN SECRETED ISOFORM).
 FT VASPLIC 1088 1255 MISSING (IN SECRETED ISOFORM).
 FT CONFLICT 2 2 T-> A (IN REF. 11).
 FT CONFLICT 134 134 P-> Q (IN REF. 9).
 FT CONFLICT 154 154 P-> Q (IN REF. 9).
 FT CONFLICT 1021 1021 S-> T (IN REF. 3).
 FT CONFLICT 1251 1251 A-> T (IN REF. 3).
 SQ SEQUENCE 1255 AA; 122072 MW; 5E28DFC4DE7D9A82 CRC64;

Query Match 12.0%; Score 651; DB 1; Length 1255;
 Best Local Similarity 27.8%; Pred. No. 1.3e-19;
 Matches 295; Conservative 72; Mismatches 460; Indels 234; Gaps 51;

DB 75 TSLVANKETTVETKETTNNKOTSDGKKTSAKTSQSTIEKSKAD-LAPTSKIYAKPT 133
 16 TLTIV-----VTGSHASSTPGGKETSATORSSSTENNAVSMSSVLSHS 65
 QY 134 P-KAETTKGP--ALTTKEP-----TPTTKPEASTPTKEPTPTT--K 173
 DB 66 PGGSGSTTGQDVTLLAPATEPAGSAAITGQDVTSPVTPRALGSTITPAHGVTSAPDK 125
 QY 174 SAPTPKEAPPTTKSAPTPKEAPPTTKP-----APTPKEAPPTTKPEAPPTTK 227
 DB 126 PABGSTAPRAHGVTSAPDT--RPAGSTAPRAHGVTSAPDT--RPAGSTAPRAHGV 179
 QY 228 SAPTPKEAPPTP-----KRAPTTKEP-----APTKEPTPTTP----- 265
 DB 180 SAPDTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGV 239

QY 266 ----KEAPTTKEP-----APTPKEAPPTAK----KRAPTTKEAPPTTK----E 306
 DB 240 SAPDTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGV 299
 QY 307 PAPTPKEAPPTTKPEAPPTT---TKSAPTTKEAPPTTKSAPTPKEAPPTTKED-- 361
 DB 300 SAPDTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAH 356
 QY 362 ----APTPKEAPPTTK-----KRAPTTKEAPPTTK-----EAPTTTKRAPTAKEP 409
 DB 357 GVTSAADTRPAPAGSTAPRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAPAG 416
 QY 410 APTPKEAPPTTKPLPTTPEKLAAPTPPEKAPPTTKPEAPPTTKPEAPPTTKPEAPPT 469
 DB 417 GVTSAADTRPAPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGVTSAPDTRPAGSTA 472
 QY 470 PKA----AAPNTPKEAPPTTK-----EPAPTTKEAPPTTKETAPTTKGAPTTKEP 521
 DB 473 PRAHGVTSAPDTRPAPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGVTSAPDTRPAP 528
 QY 522 APTP-----KRAPKEAPTTKEPTSTISDKAP--TPPKGAPTTKEAPPT 569
 DB 529 GSTAPRAHGVTSAPDTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGVTS--APD 583
 QY 570 TPKEAPPTTKG-----TAPTTKEAPPTTK-----KRAPKEAPTTTKGPTSTT 615
 DB 584 TRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGVTSAP 642
 QY 616 SDKPAP--TPK-----ETAPTTKEAPPTTK-----KRAPPTTEPTPTTSEVS 659
 DB 643 DTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGVTSAPDTRPAGS--TAPRAHGVTS 700
 QY 660 TPTTKKEPT-----IHKSPESTPELSAPPTKALENSKEGVPETTKPATKPEM 712
 DB 701 ADTPRAGSTAPRAHGVTSAPDTRPAGSTAPRAHGVTSAP-----DTRPAGSTAP-- 753
 QY 713 TTTAKDTERLRLTPEPT-----TAAPKMKETATTEKTESKITATTTGVTSTT 765
 DB 754 --PANGVTSAPDTRPAGSTAPRAHGVTSAP-----DTRPAGSTAPRAHGVTSAPDTRPA 807
 QY 766 TODTPP--FKITLTKTTLLAPKVTTKKTIYTT--EIMNKPETAKPRADRANSATTPK 821
 DB 808 PGSTAPRAHGVTSAPDTRPAGSTAPRAHGVTSAPDTRPAGSTAPRAHGVTSAPDTRPA 867
 QY 822 POKPTKAPKPTSTKKPKTPMVRKPTTPPRKMTSTMPDELNPTSRIAEMALQTTTBP 881
 DB 868 P--GSTAPRAHGVTSAPDTRP--APGSTAPRAHGVTSAPDTRPA-----PG 909
 QY 882 QT--PNSKIVEVNPKSEDAGAGETPRHMLLRPHVMEVTPDMDYLRVYNOGIINPML 940
 DB 910 STAPRAHGVTSAPDTRPAGSTAPRAH-----GVTSAADNRBALGSTA--PFV 955
 QY 941 SPEFTICNGPYDGLTTLRNGTLVAFRGHYFMMLSPSPSPS 981
 DB 956 HNVTSASGSASASTLVHNGTSABATTTTPASKSTPFSPS 996
 RESULT 5
 FPL_MYTED STANDARD; PRT; 875 AA.
 AC Q25460;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ADHESIVE PLAQUE MATRIX PROTEIN (POLYHENOGLIC ADHESIVE PROTEIN) (FOOT
 DE PROTEIN 1) (MEFPI) (FRAGMENT).
 GN FPL.
 OS Mytilus edulis (Blue mussel).
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
 OC Mytiloidea; Mytilidae; Mytilus.
 NCBI_TaxID=6550;
 RN [1]

RP SEQUENCE FROM N.A.
 RA FILIPINA D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.,
 RT "Structural and functional repetition in a marine mussel adhesive
 protein.",
 RL Biotechnol. Prog. 6:171-177(1990).
 RN 121
 RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.
 RX MEDLINE=83135732; PubMed=6298211;
 RA Waite J.H.;
 RT "Evidence for a repeating 3,4-dihydroxyphenylalanine- and
 hydroxyproline-containing decapeptide in the adhesive protein of the
 mussel, Mytilus edulis L.";
 RL J. Biol. Chem. 258:2911-2915(1983).
 CC -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
 PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
 ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
 CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
 CC -1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
 CC -1- PPM: THE DECAPEPTIDE A-K-P-S-Y-P-T-Y-K IS POST-TRANSLATIONALLY
 CC MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE
 CC HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE
 CC (DOPA) DERIVED FROM TYROSINE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X54422; CAA38294.1; -
 DR Interpro: IPR002964; Adhesive_plaq.
 DR Interpro: IPR002965; P-rich_extensn.
 DR PRINTS: PRO1216; ADHESIVE1.
 DR PRINTS: PRO1217; PRICHEXTENS.
 KW Repeat; Hydroxylation.
 FT NON_TER 1
 FT DOMAIN 67 870
 FT SEQUENCE 875 AA; 100412 MW; 6EA85312748CAACE CRC64;
 SO
 Query Match 10.3%; Score 556.5; DB 1; Length 875;
 Best Local Similarity 28.7%; Pred. No. 5,2e-16;
 Matches 272; Conservative 106; Mismatches 375; Indels 195; Gaps 56;
 QY 42 TQHNKVSITPKITAKPINP-----RP-SLPPNSDTSKETSITVNKETTVET 87
 DB 1 TKHBPVYKPVTSYSAVYKPPYQPLKKKVDYRPTKSYPPYG-SKTVYLPLAKLSYK 59
 QY 88 KETTTTKKONTDC--KEKT---SAKETOSIEKTSAKDLAPTSLVAKPPKAEITTKG 142
 DB 60 PIKTTYNAKINPPVYKPKMTYPPYKPSYPPYKSKPYKPKITTYPPYKAKPSY- 117
 QY 143 PALTTKEPTPTTPEK-----PASTTKEPTPTTISAPTPP-----KEPATTTKSAPTT 193
 DB 118 PSSYKPKKTYPPYKPLTYPPYKPKPSYPPYKPKPSYPSKTKTYSSTKAKPSY 177
 QY 194 PKEPAPTTKEPA--PTTPEKA-----PTTKEPA--PTTTSASAPTPK--EPAPTT 240
 DB 178 P-----PTYKAKPSYPPYKAKPSYPPYKAKPYKAKPTYSYKAKPSYPPYKAKPTY 233
 QY 241 PKKPA--PTTPEKPA--PTTPEKPT--PTTPEK--APTKEAPPTTPEKPA--PTAPK 289
 DB 234 KAKSYPTTAKKPSYPTTAKKPSYPTTAKKPYKAKPYKAKPYKAKPSYPPYK 293
 QY 290 KPA--PTTPEKPA--PTTPEKPA--PTTKEPS--PTTPEKPA--PTTTSAPTTKEPA- 339
 DB 294 KPSYPTTAKKPSYPTTAKKPSYPTTAKKPSYPTTAKKPSYPTTAKKPSYPTTAKKPTY 353

QY 340 PTTTSAPTTKEPSPPTTTKEPAPTTPEKPA--PTTKEPA--PTTKEPAP 395
 DB 334 PSTYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSY 409
 QY 336 TTTKRPAPTAPEAPPTTPEKPA--PTTPEKPA--PTTPEKPA--PTTPEKPA--PTTPEKPA 455
 DB 410 PPTYKAKPTTAKK-----TYPST-YKAKPSY-----PSIKAKPSYPTTAKKAPTYYKA 457
 QY 456 EPTPTTPEEAPTTTPEKAAAPNTPEKAPPTTPEKPA--PTTPEKPA--PTTPEKPA--PTTPEKPA 509
 DB 458 KPT-----YPSYKKA-----KPSYASAKKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSY 506
 QY 510 PKGTAPTTLKEPAPTTTPEKPA--PTTPEKPA--PTTPEKPA--PTTPEKPA--PTTPEKPA--PTTPEKPA 565
 DB 507 PKLITKAPTYK--PKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSY 551
 QY 556 PAPTTPEKAPPTTPEKPA--PTTPEKPA--PTTPEKPA--PTTPEKPA--PTTPEKPA--PTTPEKPA 617
 DB 552 PATYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSY 611
 QY 618 KPAPTPEETAPTTTKEPA--PTTKEPA--PTT-----PETPPTTSEVSTPTTKE----- 666
 DB 612 KAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSY 667
 QY 667 PTTIKSPDESPTPELAPPTTPEKPA--PTTPEKPA--PTTPEKPA--PTTPEKPA--PTTPEKPA--PTTPEKPA 715
 DB 668 PPT-YKAPPSYPTTAKKAPTYYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSY 726
 QY 716 AKDPTTERTDPTTPTTPTTAAAPMTKETATTTTEKTESKITATTTTQVSTTTQDPTTPTT 775
 DB 727 YKAKPTTAKAPTYPPSYKAKP--TYKAKPTTPTTPTTAKP-----PSYPPYKPKPSYPP 777
 QY 776 TLKTTTAPKVTYTTTKTTTTEIMAKPEETAKPKRAATNSKATTTPKPKAPKAPK---P 832
 DB 778 TYKSSISYPSYKPKKTYPPPT---YKPKLTYPPYK-----PKPSYPSYKPKKITYPP 826
 QY 833 TSTKPKTTPRVKPKKTYPP-----RKMTSTMPELNPTSR 868
 DB 827 STYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSY 874
 RESULT 6
 ZAN_HUMAN STANDARD; PRT; 2700 AA.
 AC Q9Y493; 000218;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ZONADHESIN (FRAGMENT).
 GN ZAN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 1-2379 FROM N.A.
 RX MEDLINE=99018118; PubMed=9799793;
 RA Glockner G., Scherer S., Schatteroy R., Boright A., Weber J.,
 RA Tsai L.C., Rosenthal A.;
 RT "Large-scale sequencing of two regions in human chromosome 7q22:
 RT analysis of 650 kb of genomic sequence around the Epo and CUTL1 loci
 RT reveals 17 genes.";
 RL Genome Res. 8:1060-1073(1998).
 RN [2]
 RP SEQUENCE OF 2338-2700 FROM N.A.
 RX TISSUE-TESTIS;
 RX MEDLINE=97271566; PubMed=9126492;
 RA Gao Z., Harumi T., Gathers D.L.;
 RT "Chromosome localization of the mouse zonadhesin gene and the human
 RT zonadhesin gene (ZAN).";
 RL Genomics 41:119-122(1997).
 CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA

OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR SIGNALING.

-1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.

-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).

-1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.

-1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE ZONA PELLUCIDA.

-1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS, THE MAM-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF SPERMATOZOEA OR PROMOTING ADHESION TO THE OVUDUCTAL ISTHMUS.

-1- DOMAIN: THE WMPD DOMAINS 2 AND 3 MAY MEDIATE COVALENT OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MU2).

-1- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.

-1- SIMILARITY: CONTAINS 4.5 WMPD DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL: AF053356; AAC78790.1; -

EMBL: U83191; AAC51208.1; -

EMBL: 602372; -

InterPro: IPR000561; EGF-like.

InterPro: IPR000998; MAM.

InterPro: IPR002965; P-rich_extensn.

InterPro: IPR002919; TIL.

InterPro: IPR003328; TILA.

InterPro: IPR001007; WWC.

InterPro: IPR001846; Wvd.

Pfam: PF00629; MAM; 3.

Pfam: PF01826; TIL; 4.

Pfam: PF02345; TILA; 4.

Pfam: PF00094; Wvd; 4.

PRINTS: PR01217; PRICHTEXTENS.

SMART: SM00137; MAM; 2.

SMART: SM00214; WWC; 1.

SMART: SM0011; WWC_def; 3.

PROSITE: PS01186; EGF_2; 3.

PROSITE: PS00740; MAM_1; 1.

PROSITE: PS50060; MAM_2; 4.

Glycoprotein; Transmembrane; Cell adhesion; Repeat.

NON_TER

DOMAIN 1

DOMAIN 109

DOMAIN 112 136

DOMAIN 161 326

DOMAIN 322 446

DOMAIN 483 951

DOMAIN 953 1065

DOMAIN 1066 1454

DOMAIN 1455 1861

DOMAIN 1862 2292

DOMAIN 2293 2684

DOMAIN 7

DOMAIN 74 74

CARBOHYD 403 403

CARBOHYD 1023 1023

CARBOHYD 1099 1099

CARBOHYD 1618 1618

CARBOHYD 1737 1737

CARBOHYD 1832 1832

CARBOHYD 1878 2136

CARBOHYD 2136 2505

CARBOHYD 2505 2379

CONFLICT 2374 2379

NON_TER 2700 2700

SEQUENCE 2700 AA; 293013 MW; 80E60CC0B12277B1 CRC64;

Query Match 10.3%; Score 555.5; DB 1; Length 2700;

Best Local Similarity 32.6%; Pred. No. 14e-15;

Matches 199; Conservative 64; Mismatches 240; Indels 107; Gaps 32;

58 PINPRESLPPNSDTSKETSJLVNKEETVETKEETTNKOTSDCKEKTYSARETOSIEKT 117

1: ||||| 1: ||||| 1: ||||| 1: ||||| 1: |||||

445 PKVLPDELPPVSPVSS-----TGPSSETTGLTENTPSTK-----KPTVSIKAP 487

118 SAKDIAPTSKVLAKPTPKAETTTGDPALTTPKEEPTTPPKEPASITPKAPT-----PTII 172

488 SYTTEKPT-----VPKEKPTIPEKPTISTEK-----PTISEKPNMSEKPTIPEKPTII 539

173 KSAPTPPKEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 231

540 TEKPTIPE-KPTIPEKPTISTEKPTVPLE--PTTPEETTTVEEVIPIPEKTSIT 596

232 TPKEPAPTTPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 291

597 --EKPSIPEK--PTISMETITISEKPTICEKPTIPEK--PTIPEKTSIPEK 647

292 APPTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 351

648 -PTTPTE-KPTIPEKPTISTEKPTIPEK-PTIPEKPTIPEKPTIPEK--EKTPTPT 701

352 EPEPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 406

702 EKTPTISTEK--PTTPTEETITISEKTSIMEKTTLTTEETITISVEETITISTEKLTPM-- 757

407 KEPAPTTTPKETAPT-----TPKKLPTTPKKA-----PTTPKEPAPTTTPEELAPTTPEE 456

758 EKPTISTEKTPTIPEKPTISTEKLTIPT-EKLTIPKEKPTIPEETITISTEKT--TIPE 814

457 PTTPTEEPAPTTTPKAAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 516

815 KPTIPEKPTISTEK-----PTIPE-KPTIPE-----ETITISTEKLTIPT 855

517 TLKEPAPTTTPKKAPKELAPTTTKEPTST-----SDKRAPTTPKGATTPPKKEPAPT 570

856 --EKPTIPEKLTPTIPEKPTISTEKLTIPEKLTPTIPEKLTPTIPEKLTPTIPEKLTPTIPEK 911

571 PKEPAPTTPKGATPTTKEPAPTTTTPKKAPKELAPTTTSGPTSTSDKRAPTTPKET--A 628

912 PPHSPPTA-TGLAALVMSHPASTPMTSV--ILGTTTTSRSTGMSCP-PNARYSCAC 966

629 PTPPKEPAPT 638

967 PASCKSPRS 976

RESULT 7

EXTN_TOBAC STANDARD; PRT; 620 AA.

AC P13963;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 01-MAR-1992 (Rel. 21, Last annotation update)

DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).

GN HRPNT3.

OS Nicotiana tabacum (Common tobacco)

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; easterids I; Solanales; Solanaceae; Nicotiana.

OX NCBI_TaxID=4097;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. XANTHI; TISSUE=Leaf;

RX MEDLINE=90128263; PubMed=2612909;

RA Keller B., Lamb C.J.;

RT *Specific expression of a novel cell wall hydroxyproline-rich

RT glycoprotein gene in lateral root initiation.*;

RL Genes Dev. 3:1639-1646(1989).

```

CC -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
CC MAIN ROOT.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X13885; CAA32090.1; -.
CC DR PIR: S06733; S06733.
CC KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
CC Hydroxylation.
CC FT SIGNAL 1 ?
CC FT CHAIN ? 620 EXTENSIN.
CC FT REPEAT 70 73 H-A-P-P.
CC FT REPEAT 148 151 H-A-P-P.
CC FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
CC FT REPEAT 229 235 1.
CC FT REPEAT 236 242 2.
CC FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.
CC FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
CC SO SEQUENCE 620 AA; 65406 MW; 641DD278AB28524 CRC64;

Query Match 10.2%; Score 551; DB 1; Length 620;
Best Local Similarity 27.2%; Pred. No. 6,6e-16;
Matches 170; Conservative 64; Mismatches 304; Indels 88; Gaps 24;

OY 136 AETTK-GPALTTP--KEPRTT---PKEPASTPKK---PPTTTKSAPTTPKKEAP 184
DB 24 AEAATQVGGYLPRLVTSQPPSSIGLSPSPAPTTPPSRGHVSP---RHAPRRHAYPKE 80
OY 185 TTTTSAPTPKREP-----APTTKKEAPRT---PKKEAPTTTKEAPTTTTSATTTTKE 235
DB 81 SHGLPSPVGGRRHNGHLPSPGRRNPSPVISPSPHPSYGAPRPSHGHLPSHGQR 140
OY 236 PAPTPPKKRAPT---TKKEAPTTKEPTTPPKKEAPTTK--EPAPTTKEAPATAKK 290
DB 141 P-----PSSHGHPRRSGGHTPRRCONPRSHKRRSPSRGNGHRRPTTYAQRPPTYIS 193
OY 291 PAPTPKEAPTTKKEAPTTTKEPSPTPK--EPAPTTKSAPTTTKKEAPTTTTS-- 345
DB 194 PSPGV--QRPPT--YSPRRPTNVQTPSPRSRGHQRPRTNRHAPRTNRHAPRTNQRPL 249
OY 346 ---APTPKESPTTKEAPRTTTPKKEAPTTKKEAPTTKKEAPRT---TKKEAPTTT 399
DB 250 RHLPSPSRKQORPTYSPPRAYAQSPRTYSPPRTYSPPRTYSPPRTYSPPRAYSPSP 309
OY 400 KPAAT---APKERPAPTTKKEAPTTTKKL--TPTT-----EKLAPTTKEAPATPPEL 449
DB 310 PRTPTPSPSPRAYSPRTYSPPRTYLPRLPSSPTYSPPRVYSPPRTYSPPRTYPL 369
OY 450 APTPPEERTPTTPEAPATTPKKAAP---NTPKKEAPTTKKEAPATTPKKEAPTTKETA 506
DB 370 PPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 427
OY 507 PTTTKGAPTTLKKEAPTTKKEAPTTKKEAPTTTKEPTSTTDKRAPTTTKEGAPTTTKEP 566
DB 428 P-----LPPTYSPPRAYSPSPRTYSPPRTYSPPRAYAQSPRPRTYSPPRAYSP 481
OY 567 APTPKKEAPATTPKGAATTTKKEAPATTPKKEAPKELAPTTTGTSTSTSKAPATTPKE 626
DB 482 PPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 538
OY 627 TAPTPKKEAPAT--TPKKRAPATPPEPTTSEVSTPTTKEPTTIHNSPDESTPELSAE 684

```

```

DB 539 PPPRQIHSPRRPQPRPTTPYGGPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 598
OY 685 PTPKALNSPKRGVPTTKTPATKP 710
DB 599 PSP-----PTTYSPPSPPP 612

RESULT 8
NFH_MOUSE
ID NFH_MOUSE STANDARD: PRT: 1087 AA.
AC P19246; Q61959;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
GN NFH OR NFH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RX MEDLINE=89121513; PubMed=3220257;
RA Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosved F.,
RA Mushynski W.;
RT "Sequence and structure of the mouse gene coding for the largest
RT neurofilament subunit."
RL Gene 68:307-314(1988).
RN [2]
RX MEDLINE=89089138; PubMed=3145094;
RA Sheidman P.S., Carden M.J., Lees J.F., Lazarini R.A.;
RT "The structure of the largest murine neurofilament protein (NF-H) as
RT revealed by cDNA and genomic sequences."
RL Brain Res. 464:217-231(1988).
RN [3]
RX SEQUENCE FROM N.A.
RN [4]
RX SEQUENCE FROM N.A.
RN [5]
RX STRAIN-SWISS WEBSTER; TISSUE=Brain;
RA Carden M.J.;
RT Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC OBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M24496; AAA39813.1; -.
CC DR EMBL: M23349; AAA39813.1; JOINED.
CC DR EMBL: M24494; AAA39813.1; JOINED.
CC DR EMBL: M24495; AAA39813.1; JOINED.
CC DR EMBL: M35131; AAA39809.1; ALT_FRAME.
CC EMBL: 231012; CAA83229.1; -.

```



```

0X NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92178992; PubMed=1542581;
RA Witzgerich M., Acker J., Vicaire S., Vigneron M., Keding C.;
RL "Complete sequence of the human RNA polymerase II largest subunit.";
RN Nucleic Acids Res. 20:910-910(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95347616; PubMed=7622068;
RA Mita K., Tsuji H., Morimyo M., Takahashi E., Nenoi M.,
RA Ichimura S., Yamauchi M., Hongo E., Hayashi A.;
RL "The human gene encoding the largest subunit of RNA polymerase II.";
RN Gene 159:285-286(1995).
CC -I- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -I- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -I- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
CC -I- SUBCELLULAR LOCATION: NUCLEAR.
CC -I- PWM: THE TANDDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC THE PHOSPHORYLATION ACTIVATES POL2.
CC -I- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND TRNA GENES.
CC -I- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: X63564; CAA45125.1; -.
DR EMBL: X74874; CAA52862.1; -.
DR EMBL: X74873; CAA52862.1; JOINED.
DR EMBL: X74872; CAA52862.1; JOINED.
DR EMBL: X74871; CAA52862.1; JOINED.
DR EMBL: X74870; CAA52862.1; JOINED.
DR PIR: S21054; S21054.
DR MIM: 180660; -.
DR InterPro: IPR000684; RNA_polII_repeat.
DR InterPro: IPR000722; RNA_pol_A.
DR InterPro: IPR002879; RNA_pol_A2.
DR Pfam: PF00623; RNA_pol_A; 1.
DR Pfam: PF01854; RNA_pol_A; 1.
DR PROSITE: PS00115; RNA_POL_II_REPEAT; 43.
KW Transferrase; DNA-directed RNA polymerase; transcription; zinc; Repeat;
KW DNA-binding; Nuclear protein; Phosphorylation; zinc-finger.
FT ZN_FING 71 87
FT DOMAIN 1590 1958 C2H2-TYPE (POTENTIAL).
FT CONFLICT 1067 1067 W -> L (IN REF. 2).
FT CONFLICT 1449 1449 D -> Y (IN REF. 2).
SQ SEQUENCE 1970 AA; 217205 MW; 6876FC23692A657E CRC64;

Query Match 9.3%; Score 503.5; DB 1; Length 1970;
Best Local Similarity 33.3%; Pred. No. 1,2e-13;
Matches 181; Conservative 101; Mismatches 167; Indels 95; Gaps 47

0Y 151 PPTPTTKEPASTPKKE-PPTTTIKSAPTTKEPAPTITTKSAPTTKE-----PAPTITKE 204
0Y 1507 PSMWGGISPMATPMWNGCAPAYGAWSPSVSGMTPGCAAGFSPASASGFSBGSPAWS 1566
0Y 205 PAPTITKEPAPTITTKEPAPTITTKSAPTTKEPAPTITTKKRAPITPKPE--APTITPKPPT 262
0Y 1567 PPTGSPSPSPSPSSPIYIPSGGAWSPSPYSIPA-YEPRSPGGITPPQSPSPSPSPSPSP 1624
0Y 263 TTPKEPAPTITKEPAPTITTKEPAPTAPKPAAPTITTKEPAPTITTKEPAPTITTKESPTTKE 322

```


01 677 STEP 680
 Db 1966 SDEE 1969
 RESULT 15
 ID NFH_RAT
 AC p1684; STANDARD; PRT; 831 AA.
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).
 DE NEFH OR NFH.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CX NCBI_TaxID=10116.
 [1]
 SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=89065087; PubMed=3143606;
 RA Bren K.C., Robinson P.A., Wilson D., Anderson B.H.;
 RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
 RT Identification of putative phosphorylation sites.";
 RL FEBS Lett. 241:213-218(1988).
 [2]
 SEQUENCE OF 37-831 FROM N.A.
 RP MEDLINE=88309090; PubMed=2457365;
 RX Dautigny A., Pham-Dinh D., Rousset C., Felix J.M., Nussbaum J.L.,
 RA Jolles P.;
 RT "Type large neurofilament subunit (NF-H) of the rat: cDNA cloning and
 RT in situ detection.";
 RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
 [3]
 SEQUENCE OF 1-89 AND 243-313 FROM N.A.
 RP MEDLINE=87080760; PubMed=2878828;
 RX Robinson P.A., Wilson D., Anderson B.H.;
 RA "Isolation of a cDNA for the rat heavy neurofilament polypeptide
 RA (NF-H).";
 RT FEBS Lett. 209:203-205(1986).
 [4]
 SEQUENCE OF 318-831 FROM N.A.
 RP MEDLINE=89184647; PubMed=2928342;
 RX Lieberburg I., Spinner N., Snyder S., Anderson J., Goldhaber D.,
 RA Smolowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;
 RT "Cloning of a cDNA encoding the rat high molecular weight
 RT neurofilament peptide (NF-H): development and tissue expression in
 RT the rat, and mapping of its human homologue to chromosomes 1 and
 RT 22.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
 -I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 -I- AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 -I- NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 -I- SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 -I- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
 -I- PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 -I- THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 -I- INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 -I- OF AXONAL CALIBER.
 -I- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 -I- OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 -I- LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 -I- COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 -I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY
 -I- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
 ONWARD AND IS LONGER DUE TO A FRAMESHIFT.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR	EMBL: M219644: AAA41695..1; .	
DR	EMBL: J04517: AAA41692.1; .	
DR	PIR: A30796; A30796.	
DR	PIR: A25649; A25649.	
DR	PIR: B25649; B25649.	
DR	PIR: S02003; S02003.	
DR	InterPro: IPR001664; IF.	
DR	Pfam: PF00038; filament; 1.	
DR	ProSite; PS00226; IF; 1.	
KW	Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.	
FT	NON_TER	1
FT	DOMAIN	1
FT	CONFLICT	276 641
FT	CONFLICT	164 164
FT	CONFLICT	185 185
FT	CONFLICT	193 193
FT	CONFLICT	199 199
FT	CONFLICT	346 346
FT	CONFLICT	373 373
FT	CONFLICT	482 482
FT	CONFLICT	485 485
FT	CONFLICT	570 571
FT	CONFLICT	591 591
FT	CONFLICT	727 727
FT	CONFLICT	757 759
FT	CONFLICT	769 769
FT	CONFLICT	775 775
SQ	SEQUENCE	831 AA: 89486 MW: 1B0973C3f13EF768 CRC64:

Query Match	9.28;	Score 497;	DB 1;	Length 831;
Best Local Similarity	27.68;	Pred. No. 1.1e-13;		
Matches 195;	Conservative 59;	Mismatches 307;	Indels 146;	Gaps 35

```

0Y 37 PDSTJQHNKXSPSPITIAKPIINPABSLPDNDSTKESLFWNKRTTVEIKETTNNQ 96
Db 197 PSMST--HKNVSEEEKIAYE-----KSKETVIYBEQJEBELQVNEEVEED 242
0Y 97 TSTGKEKTTSAKETOSIEKTSAKDLAPITSKYLIAPTPAETTTKGPALITTKPEP-----T 152
Db 243 KEAQGESEEAEGGEAAATTS-----PAAEEAASP-----EKETKSPVEEAKSPEAKS 293
0Y 153 PTTTKPEPA-STPKEEPTPTTIKSAPITTKPEPA-----PTTTSAPITTKPEPA-----PTTK 203
Db 294 PAAKAPSAEKSPAEVKSPPAVAKSPAEVKSPPALVKSPPAEKS--PAAKAPSAEVKSPPATYK 352
0Y 204 EP-----APITTKPEPA-----PTTKKEPA-----PTTTSAPITTKPEPA-----PTTKKPA-- 245
Db 353 SPEAKSPAEKSPALVKSPPAEKSPAEKSPASVKS--PEAKSPAEKSPAEVKSPPAV 411
0Y 246 --PTTKPEPA-----PTTKPEP-----PTTKKEPAPTTKPEP-----APTTKPEPA-----PTA 287
Db 412 KSPVEAKSPPEVKSPPVTVSPAEKSPVEVKSPPA-SYKSPSEKSPAGAKSPAEKSPV 470
0Y 288 PKRPAPITTKPEPAPTTPKEPAPTTKKEPSPTTKPEPA---PTTKTS-----APTTKEP 338
Db 471 AKSPAAEA--KSPAGAKPPAPAEKSPAEKSPAEKSPAEKSPAEKSPAEVKSPPAEKSP 538
0Y 339 APITTKS-----APITTKESPITTKPEPA---PTTKPEPAPTTPKPPAPTTPKPEPAPT 389
Db 529 VKEGANSILAEBKSPAEKSPVEKEIKPDAEVKSPEKAKSPMKREAKSPPEAKTILDVASPE 588
0Y 390 PKPEPAPTTPKPPAPTAPEKEPAPTTPKEPAPTTPKLIPTTPEKLIAPTTPPEKPAFTTPEEL 449
Db 589 AKPPAKKEAKRPADIRSEQVKSAPKEEAKSPEKEETRT--EKVAPKPEEVKSPV--EEV 644
0Y 450 APITTPPEPTPTTPEEPAPTTPKAAAPNTPEPAPTPTTKKEPAPTTPKPPAPTTPKETAPT 509

```

Db 645 ---KAKEPPKVEEKKTPATPKTEVYESKDKDEAPKEAOKP-KAEKEPLTEKPKDSPGEA 700
OY 510 PKGTAPTTLKEBPAPTTPKKPAKELAPTTTKEPTSTTSKRPAPTTPKGTAPTTPKBPAPY 569
Db 701 KKEEA---KEKKAAPPEETPAKLGVEEAKPKKAKADAKA-----KEPSKP 744
OY 570 TPKEBPAPTTPKGTAPTTLKEBPAPTTPKKPAKELAPTTTGTSTTSKRPAPTTPKETAP 629
Db 745 SEKE---KPK-----KEEVPAPPEKKDTKE-----EKTESKKKEEKPKMEAK 784
OY 630 TTPKEBPAPTTPKKPAKELAPTTTKEPTSTTSKRPAPTTPKGTAPTTTKEPTTIHKSDE 676
Db 785 AKED-----KGLPQEPSPKPKTEKAESSTDOKDSOPSEKAPED 824

Search completed: April 26, 2002, 16:20:03
Job time: 388 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:17:20 ; Search time 49.78 seconds

(without alignments)
457.932 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1212

Perfect score: 5416
Sequence: 1 VKDNKKNRTKKRPPVY.....SPIDVFTRCNCEGKTFPEK 1013

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/pcodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/pcodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/pcodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/pcodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/pcodata/2/1aa/PCRTUS_COMB.pep:*
6: /cgn2_6/pcodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	522	9.6	805	4	US-09-103-429A-4
2	508	9.4	1837	4	US-08-928-361B-5
3	506.5	9.4	744	6	5202236-25
4	498.5	9.2	786	4	US-09-103-429A-3
5	489	9.0	1721	3	US-08-700-651-5
6	489	9.0	1721	3	US-08-928-361B-6
7	488.5	9.0	826	1	US-07-638-431-2
8	488.5	9.0	826	5	PCF-US92-00018-2
9	476.5	8.8	652	6	5202236-13
10	452	8.3	960	4	US-09-219-849-5
11	424.5	7.8	1867	2	US-08-479-537A-5
12	424.5	7.8	1867	4	US-09-083-116-5
13	424.5	7.8	2035	2	US-08-479-537A-2
14	424.5	7.8	2035	4	US-09-083-116-2
15	419.5	7.7	2476	2	US-08-276-967-2
16	419	7.7	1185	4	US-09-041-886-23
17	417	7.7	829	1	US-08-642-255-132
18	417	7.7	829	1	US-08-397-633A-53
19	417	7.7	837	1	US-08-175-155-68
20	417	7.7	837	1	US-08-477-509B-103
21	417	7.7	837	1	US-08-642-255-101
22	417	7.7	837	2	US-08-707-237A-75
23	417	7.7	837	3	US-08-482-085B-103
24	417	7.7	897	1	US-08-397-633A-50
25	413.5	7.6	907	3	US-08-783-774-2
26	413.5	7.6	907	5	PCF-US95-04611A-19
27	404.5	7.5	408	1	US-07-609-716-65

28	404.5	7.5	408	4	US-08-475-411A-65	Sequence 65, App1
29	404.5	7.5	408	4	US-08-478-029A-65	Sequence 65, App1
30	398.5	7.4	682	1	US-08-642-255-126	Sequence 126, App
31	398.5	7.4	682	1	US-08-397-633A-36	Sequence 36, App1
32	381	7.0	1848	4	PCF-US95-791-6	Sequence 6, App1
33	381	7.0	1848	5	PCF-US95-10611A-6	Sequence 6, App1
34	376	6.9	960	4	US-09-219-849-6	Sequence 6, App1
35	374.5	6.9	1537	3	US-08-325-267A-2	Sequence 2, App1
36	373	6.9	1231	3	US-08-904-263A-4	Sequence 4, App1
37	371	6.9	761	2	US-08-707-237A-84	Sequence 84, App1
38	371	6.9	762	1	US-08-642-255-114	Sequence 114, App
39	371	6.9	762	1	US-08-397-633A-26	Sequence 26, App1
40	370.5	6.8	1064	1	US-08-642-255-62	Sequence 62, App1
41	368	6.8	762	1	US-08-320-559-28	Sequence 28, App1
42	367.5	6.8	762	1	US-08-320-559-28	Sequence 28, App1
43	367.5	6.8	1187	3	US-08-545-860D-28	Sequence 28, App1
44	367.5	6.8	1187	5	PCF-US94-04496-28	Sequence 28, App1
45	367.5	6.8	1187	5	PCF-US94-04496-28	Sequence 28, App1

ALIGNMENTS

RESULT 1
US-09-103-429A-4
; Sequence 4, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang, Ping
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Nuc1n
; TITLE OF INVENTION: CNNA and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Plunisi & Michaels, P.C.
; STREET: 118 No. 6187558Ch T10ga
; CITY: Itasca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,429A
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-2000
; TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 805 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; TISSUE TYPE: peritrophic membrane
; US-09-103-429A-4

Query Match 9.6%; Score 522; DB 4; Length 805;
Best Local Similarity 27.8%; Pred. No. 2.4e-26;

```

CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verity, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1837 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-5

```



```

QY 412 -----TTPKETA-PTTP-----KKL 425
Db 842 AGQADTSLNLFVQTHKSGLEIDPMVGLPEDPKSGNLVHPYTNQTMGLSGLSYLAANKL 901
QY 426 T-----PTPEKLAPTTPE-----439
Db 902 TVDTDEYGLPDLTLTGTPLEDSVLIPNPETGELFDLISDEINNGTACIVSGISASES 961
QY 440 ----KPAFTPEE-----LAPTT-----DEEPTPTTP--EPPAPTTPKAA 474
Db 962 LLSQKAPIDPATNMVGEFGLNIPATGVMIPLSGLSGEOTPPSPLEDDGIIPEEVA 1021
QY 475 PNTPEAPPTPKPEAPPTTPE-----PA 498
Db 1022 ANADFKLSIPPSVPEISPEKDKIDSISELMYDIESGRLIGQVSKREIPGSIAGDLNPI 1081
QY 499 PTPPEAPPTPKGAPPTTKEAPATTPPKAPKAPKELATTTKEPTSTTSKAPATTKGT 558
Db 1082 MKTPQTOSVYKGRIDPTT---GLPFPD--PTGHLINFTNNNTWDSSPAGAYKAVANGSI 1136
QY 559 APPTT---PKPEAPPTPKPEAPPTPKGAPPTTKEAPATTP-----KKPAK-- 601
Db 1137 KTDVNYGLPDEITGLPKDPVSDIPFNSTGELVDSGKINNYTAGIYSGKGLPE 1196
QY 602 ----ELAPTTKGP-----TST-----SDKAPTTPKETAPTTKEPA-- 636
Db 1197 DENGNLDPSTKLDIDGNOLNBPETNSVSGSTGSKRPKGPVNGGVGVPDEEAKDQ 1256
QY 637 -----PTTPKKAPPTTPEPTTSEVSTPTTKE--PTTIKKS--PDESPPEL 681
Db 1257 ADKGDGLIAPPTNSINKDPVTNQTNSNTGNIINPETKVIPLSGLSLNPSEFNPQ 1316
QY 682 SAETPKALE---NSPKPEGVPTTKTPAATKPEKTTAKKTERDLRTTPETTAARKM 738
Db 1317 TDEITGKPVDDVGLPYDPSTGELIDPATKLPVGSVAGDEILTEVINTTDEVTLGP-I 1375
QY 739 TKETATTEKTESKITATTQVSTTQDPTTPKITTLKTTTLAPKVTITTKTITITE- 797
Db 1376 DLEGLRDP-----VSGRLPDLNGLVD-----PSNKKPITPGSHS 1411
QY 798 --LNKKEETAKKPRDAYSNAKATTPKPKP--TKAP--KKPTSKKPKTPRVRKPKT-- 849
Db 1412 GFINGTSGEOSHEDKPDSTG---KPLDNTGGLPDEDSGSLINPETGKLDGSHSGT 1465
QY 850 TTPPKKSTMPLENPISRLAEAMLOTTTPRNOTPNSKLVENPK--SEDA----- 898
Db 1466 MPVPKAGQGENGIMPEQILEAL-----NKLPTSNENVISPRSSSDAVDPRTNTWM 1518
QY 899 -----GGAEGETPHML-----RPHVPEVTTPMDVTPRPNQGI 936
Db 1519 NKISGQTFQVDAKKTIPGSAASYIHATLGPPTQDPTTGLP--SDPSTGLPFIPEGNVLV 1576
QY 937 NPLMSD-----ETNIC-----NKPVDGLTTL--RNGTLV-AF 966
Db 1577 DPQTGQIKGSVYVSLYKKNIVTEAAYGLPVDPKTGFPIIDISYLPFAKNGELIDPI 1636
QY 967 RGHYF 971
Db 1637 SGKYF 1641

```

```

RESULT 3
5202236-25
Patent No. 5202236
APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSSBERG,
SUSAN L.; MCCANDLISH, RUSS WET, TENA; FILIPULA, DAVID
TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
PROTEIN
NUMBER OF SEQUENCES: 39
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/528,762
FILING DATE: 25-MAY-1990
APPLICATION NUMBER: 82,456

```

```

FILING DATE: 07-AUG-1987
APPLICATION NUMBER: 933,945
FILING DATE: 24-NOV-1986
APPLICATION NUMBER: 650,128
FILING DATE: 13-SEP-1984
SEQ ID NO: 25
LENGTH: 744
5202236-25

```

```

Query Match          9.4%; Score 506.5; DB 6; Length 744;
Best Local Similarity 29.9%; Pred. No. 2.2e-25;
Matches 239; Conservative 78; Mismatches 308; Indels 175; Gaps 50;

```

```

QY 51 PKITTAAPINRPSLSPNSDTSKETSILYVNETTVERKETTNNKOTSDGKEKTSAKE 110
Db 24 PKMTYPTPKKPSYPP-----TKSKPTLY-----KKIT----- 53
QY 111 TQSIKTSAMDAPTSKVLAKPTPKAETTTKGALTTPEEPPTTPE---PASTTPE 166
Db 54 -----YPTTK--AKPS-----YPTTKPKTYPTTKPKLYVPTTKPKP 92
QY 167 PPTTKSAAPT--PKPEAPPTTKSAPTTPKEAPATTTKPEAPATTPKEAPATTTKEPA--P 223
Db 93 SYPTTKSKPTTKYKILYPTTKAKPSYPTTKPKTTPPT--YKGLTTPPTTKKASTP 151
QY 224 TTTKSAPTTPKEAPATTPKKAAPT--PKPEAPTT-----PKPEPTTPKPEAPTTKEP-- 275
Db 152 PTYKPKPSYPP--PSYTKTKTYPTTKPKLYPTTKPKPSYPPSYKPKTYPTTKPKLT 209
QY 276 -APTTPKEPA--PTAPKKA--PTTPKEPA--PTTPKEPA--PTTKKES--PTTPKEPA 324
Db 210 YPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPS 269
QY 325 -PTTTSAPTTPKEAPATTTKSAPTTPKESPTTKKEPA--PTTPKEPA--PTTPKPK 377
Db 270 YSTYKAKPSYT---YPSYTKAKPYTP---PTYKAKPSYPTTKAKPYTKAKPK 322
QY 378 A--PTTPKEPA--PTTPKKA--PTTTPKKA--PTAPKPEAPATTTKEAPATTTKEP 429
Db 323 SYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPS 382
QY 430 PEKLAFTPE--KPAFTPEELAPTTPEEPT--PTTPEEPA--PTTPKA--AAPNT-----P 478
Db 383 TYKAKPSYPTTKAKPSYPP---PTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 438
QY 479 KPEAPTTKEAPATTT---PKPEAPTTKEAPATTT--PKGAPATTLKBPATTPKPKAPK 532
Db 439 SYPTTKPKISYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 498
QY 533 ELAPTTKEPTSTSDKAPATTPKGTAPTTPEKAPATTPKPEAPATTPKGTAPTTKEPA- 591
Db 499 SYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 546
QY 592 -PTTKKAPKELAPTTTKGPT--STTSKAPATTPKGTAPTTKEAPATTPKPKPA--PT 646
Db 547 PPTTKAKPYTK-AKPYTKAKPYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPT 601
QY 647 ---TPEPTPTTSEVSTPTTKE---PTTIHKSDD-----ESTPELSAEPPT--KPALENS 693
Db 602 YKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAK 661
QY 694 PREPGVPTTKTPAA-----TKPEMTTAKDKTERDLRTTPETTAAR-----KMT 739
Db 662 PSYPP--PTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 719
QY 740 KETATTTKTESKITATTT 759
Db 720 YKAKPYPTTKAKPYPTST 739

```

```

RESULT 4
US-09-103-429A-3

```

```

Sequence 3, Application US/09103429A
Patent No. 6187558
GENERAL INFORMATION:
APPLICANT: Granados, Robert R
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
TITLE OF INVENTION: CDNA and Related Products and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Pinisti & Michaels, P.C.
STREET: 118 No. 6187558th 110ga
CITY: Itasca
STATE: NY
COUNTRY: USA
ZIP: 14850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,429A
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Michaels, Christopher A
REGISTRATION NUMBER: 34,390
REFERENCE/DOCKET NUMBER: BTI-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: (607) 256-2000
TELEFAX: (607) 256-3628
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 786 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoplusia ni
TISSUE TYPE: peritrophic membrane
US-09-103-429A-3

```

```

Query Match          9.2%; Score 498.5; DB 4; Length 786;
Best Local Similarity 29.1%; Pred. No. 7.6e-25;
Matches 200; Conservative 25; Mismatches 207; Indels 255; Gaps 32;

```

```

OY 132 PPKAETTTKGAULTTPEPTTTPKEPASTTPKEPTTTIKSAPTTPKEPAPTTKSA 191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 93 PEPPE-PTQAATT-----QAPTTTQAATT-----QAPTTTQ--ATTTQAP 135
OY 192 TPKPAPPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKAPPTPK 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 136 TTQ--APTQ--ATT-----QAPTTTQ--APTQQAATTQ--APTQ--APTQ-- 183
OY 252 PAPTPKEPTTPKEPAPTTPKEPAPTTPKAPPAKPPAPTTPKEPAPT-- 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 184 -APTQAATTITQAATTAPATTPAATTPAATPAA--TTAATTGAVAPTSAPVWPIC 241
OY 309 -----PTTKEPSPTPK- 321
DB 242 ELLPNCAPADFDILLIPDKYCNLFYQCSNGTYEQRCPEGLYNPVYQRDSPANVC 301
OY 322 ----EPAPTT----- 328
DB 302 DGEISAPVTEGNEDEDIDIGLDNCCPANEIDMLPHGNRCDKYQCVHGNLVER 361
OY 329 -----KSAPTTPKEPAPTTPKSAPTTPKEPSPPTTPKEP 361
DB 362 CGAGTHSFELQOCHIELVGLTLPGESEEVADDEA-CTGMVCTPEITENEPNCGP 420

```

```

OY 362 APT-----PKP-----APTTPKKAPPT----- 381
DB 421 ADFSIDHLLPHESDCGYLOCVHGOTIARPCPGNLIHSPATQSESPVTAQCQVFECDSD 480
OY 382 ---PREPAPTTPKEPAPTTPKPPATPAKREPAPTTPKETAPTTPKLTPTTPEKLAPTP 438
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 481 NCSTSTAPTAAPTAAPTAAPTAAPTA--APSTVVPATPPATAAPVPTT---AAPT-- 533
OY 439 EKPAPTTPELAPTTPEEPTTPTEPAPTTPKAAAPT---PKEPAPTTPKEPAPTPE 496
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 534 --PAPTAAPTAAPTAAPESPTTVP--PTAAPTAPTAAPTAAPVETIVT---SAPTAPT 587
OY 497 PAPTPKETAPTTPKGTAPTTPKEPAPTTPKPPAKELAPTTPKEPTSTSDKAPPTPK 556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 588 AAPTAAPTAAPTAAPVETIPTVTSPTTAPTAP---APTNT-----VTVVP 632
OY 557 GAAPTTPKEPAPTTPKEPAPTTPKGTAPTTPKEPAPTTPKAPKELAPTTPKGTSTTS 616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 633 TAAPT--AAPAPNTTVVPTTAAPTAPTAAPVAH-----APTNTAAVTTTS 677
OY 617 DKPAPTTPKETAPTTPKEPAPTTPKKP 643
DB 678 ---APATPDDDDIDP--PLPNDPINP 699

```

```

RESULT 5
US-08-700-651-5
Sequence 5, Application US/08700651B
Patent No. 6015882
GENERAL INFORMATION:
APPLICANT: PETERSEN, CAROLYN
APPLICANT: LEECH, JAMES
APPLICANT: NELSON, RICHARD, C.

```

```

APPLICANT: GUT, JIRI
TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
FILE REFERENCE: 480.19-4(HV)
CURRENT APPLICATION NUMBER: US/08/700,651B
CURRENT FILING DATE: 1997-08-14
EARLIER APPLICATION NUMBER: 08/415,751
EARLIER FILING DATE: 1995-04-03
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 1721
TYPE: PRT
ORGANISM: Cryptosporidium parvum
US-08-700-651-5

```

```

Query Match          9.0%; Score 489; DB 3; Length 1721;
Best Local Similarity 22.1%; Pred. No. 7.3e-24;
Matches 284; Conservative 82; Mismatches 476; Indels 442; Gaps 41;

```

```

OY 5 KKNFTKKKPPKPPVNDACSGLDNGDFKVTYTDISTONKXSTSPKITTAAPINPRPS 64
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 116 RSNFTK--TTPEANTYAGVRSN--ETKTEPSANT--NFLVDPKLT-----N 158
OY 65 LPPNSDSKSE-----TSLVANKETVETETTTNKTSTGCKEKTSAK 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 159 APCNSSENSFEGQILFMGSKVYIPYKCGVKHHTTTTTTTTTTTTTTTTTTTT 208
OY 110 ETOSIERTSADLAFTSKVLAKPTPKAETTTKGDALTTPKEPTTPKEPASTTPKEPTP 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 209 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 240
OY 170 TTISAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSA 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 300
OY 230 PTPKEPAPTTPKPPAPTTPKEPAPTTPKEPTTPTPKEPAPTTPKEPAPTTPK 289

```

[illegible]

```

US-08-928-361B-6
; Sequence 6, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIRKA
STREET: 385 Sherman Avenue, Suite 6
City: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verny, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ. ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1721 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-6

Query Match          9.0%; Score 489; DB 3; Length 1721;
Best Local Similarity 22.1%; Pred. No. 7.3e-24;
Matches    284; Conservative      82; Mismatches 476; Indels   442; Gaps    41;

QY      5 KKNRRKKKKTTPRPVVDEAGSGLDNGDFKYTTTDDITGTHOHNVTSFKITTAPINRPS 64
DB      116 RSNERTK---TLTESANTYAGVYRSN-EIKTTESANT--NFLVPDKI-----N 158
              :| | | | | | | | | | | | | | | | | | | | | | | | | |
QY      65 LPNDSJTSKE-----TSLTVNKETVTVEKETTTTTNNQTSDGKEKNTSAK 109
DB      159 APCSNSENSEPQGQIFDMGSKVIPIRYKCQGVKHHTTTTTTTTTTTTTTTTTTTT 208
              || | | | | | | | | | | | | | | | | | | | | | | | |
QY      110 ETQSIEKSNAKLAFYSKVLAKPTPKAEFTTYTGSPALTTPKEPTPTTPKEBASTPREPP 169
DB      209 -----                || | | | | | | | | | | | | | | | | | | | |
              :||| | | | | | | | | | | | | | | | | | | | | | | | |
QY      170 TTINKSAPTTPKEPAPTTKSAPTTPKEAPAATTKKEBAPTTPKEAATTKKEAPATTKSA 229
DB      241 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 300
              || | | | | | | | | | | | | | | | | | | | | | | | |
QY      230 PTTRKEBPATTPKKAPATTPKEBAPTTPKEEPTPTTPKEBAPTTPKEBAPTTPREAPTA 289
DB      301 TTTTKKPPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 350
              || | | | | | | | | | | | | | | | | | | | | | | | |
QY      290 KPAPTTPPEAPATTPKEBAPTTPKEBSPTTPKEAPAATTKSAPTTKKEBAPTTPKSAPT 349
DB      351 ----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKPTTTTTTTTTTGTATATTT 406
              || | | | | | | | | | | | | | | | | | | | | | | | |

```

```

QY 350 PKE-----PSPT----- 357
DB 407 TSETESVIRKDEMCLEKNGEAGAIYVIGKDRIGENCAFTMLPNDOTVHFRPK 466
QY 358 -----TKEP-----ATTKE 368
DB 467 VKDVNTISVRCRKGAGLEFPDRSLDFTIPVAGHNSCSIIYGVSGDKIHVSFGSKD 526
QY 369 PAPTTPKKAPPTPKP-----PAPTTPKKAPPTPKPAPTKEPA 410
DB 527 VSLIS---AIPQSELENEVYCDITAKYGAHSGYQSADPVTITTKPPTTT--TGA 580
QY 411 PTPPKETAPPTPKLPTTPPEKLAFTTPPKPAPTPEELAPTPPEEPPTTPPEEPAPTTP 470
DB 581 PGPPTTTTGSPEKPTTTTTKATTTT-----TTLNPIITTTTQKPTTTT-----TT- 627
QY 471 KAAPPTPKPAPTTPPKPAPTTPPKPAPTTPKETAFTTPPKGAPTTPKAPTTPKPA 530
DB 628 --KVPGRPIATTTTLKPIVTTTTTKATTTTTTTPPTT-----TTTKRDEKTTTTPL 680
QY 531 PK-----EIAPTTKEPTS-----TSDKPAFTTPKG-TAPTTPKEPAPTTP 570
DB 681 PDIGDIEITPIPEKMLDKYTMIIYVNSGLLSDNDEPIPSQAQIADTSMLEFVQTH 740
QY 571 PKPAPTTPKGTAPTTLKE---PAPTTPKKAPKELAPTTTGKPTTSDKPAFTTPKET 627
DB 741 KSTGLPIDPVGILPFPKSGNLVHPYNTQMSGLSVYLAKNMLVDTDEYTG--LPIDT 798
QY 628 APTTPKEPAPTTPKKAPPTTPETTPPTTSVSTPT-----TTKEPTTIHKS----- 673
DB 799 LTGLPIDVSLIFPN--PEIGELFDPISDIMNGITAGIVSGISASESLISQKSAIDPA 856
QY 674 -----PDEST-----PELSAETPPALENSK 695
DB 857 TNMVYGEFGLNPATGVMIPGELGSEOTQFSELEDGIIPEVYAAANADKFKLSIP- 915
QY 696 EPGVPTTKTAAIKPEMTTAKDKTTER-----DLRTPEPTTAAPKM 738
DB 916 -PSP-----ESTPE-----KQOKIDISELMADIESGLIGOVSKRPIPSGIADLNP 963
QY 739 TKETAFTTEKTESKITATY--TOVTSITTTODTTPFKITTLKTTTLAPKVTYTTKKTITTE 797
DB 964 IMKPTOTDSVGKPIDPTGLPENRPTGHLINFTNNNTMDSAGAKYAVANGIKITDN 1023
QY 798 IMNKP--EETAKPKD-----RATNSKATTPKQKPTKAPKKTSTKKKKTMP----- 842
DB 1024 VYGLPVEIGTGLPKDPSDIPFNSTGTGLVDPSTGKPIINNSTAGIYSGKFGLEPDEDNG 1083
QY 843 -----RVKPKTPTPKKMTS--TMPE----- 862
DB 1084 NLFPDSTNLPIDGNNOVLNPETNSTVSGSISGTTKPKPGIPYNGGVVPEBEAKQADKG 1143
QY 863 -----LNFSTIADAMLOTTTRPNQTPNSKLVENPKSEBDAGAGEGETPHMLLRPHFMP 917
DB 1144 KDGILVPTNINSINDPVTNIOYSMTGTNI---INP--ETGKVIPTGSLPSGLNPSFNTP 1197
QY 918 EYTPD-----MDYLPRVNOGIITNP-----MLSDETNINCKN----- 950
DB 1198 QOTDEITGKPVDTVYGLYDPSGTGEIIDPATKFLIPGSVAGDELITLEVANTTDEYVGLP 1257
QY 951 -----PVDGLTTLRNGTLV 964
DB 1258 IDLEGTLPDRDPVSGLPOLPNGTLV 1281

```

```

RESULT 7
US-07-638-431-2
: Sequence 2, Application US/07638431
: Patent No. 5198535
: GENERAL INFORMATION:
: APPLICANT: Hoffman, Stephen L.
: APPLICANT: Charoenvit, Yupin

```

```

APPLICANT: Hedstrom, Richard
APPLICANT: Khumsmith, Srisin
APPLICANT: Rogers IV, William O.
TITLE OF INVENTION: Protective malaria sporozoite surface protein
TITLE OF INVENTION: Immunogen and gene
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: A. David Spevack
STREET: NMRC Building 1 T-12 National Naval
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20814-5044
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/638,431
FILING DATE: 19910110
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, Avrom D.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-4033
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 826 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-638-431-2

Query Match 9.0%; Score 488.5; DB 1; Length 826;
Best Local Similarity 26.4%; Pred. No. 3.5e-24;
Matches 153; Conservative 54; Mismatches 223; Indels 149; Gaps 34;

QY 168 TPTTTSAPPTTKEPA-PPTTSAFTTKEPA-PPTTKEP-----APTTPKEP-APTTKE 220
DB 273 TPKVYDCQIPLPIYIPKIKIEKSNPEEYVNPDPNDPNPNPNPNPNPNPN 332
QY 221 PAPTTPKAPPTTKEP-APTTPKKP-APTTPKEP-APTTPKEP-TPTTPKEPAPTKEPA 276
DB 333 PNNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 392
QY 277 PTPPKPAPAPAK--KPAFTTKEPAPTTPKEPAPTTPKEPSPTTPKEPAPTTPKSAPTT 334
DB 393 RNNPKRRNPKPKPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 443
QY 335 TKEPAPTTPKSAFTTKEPSPTTKEPAPTTPKEPA-PPTPKKP-APTTPKEPAPTTPKE 392
DB 444 PKNPNP--NEPSNNKPNPN-----EPLNPNPNPNPNPNPNPNPNPNPN 492
QY 393 PAPTTPKAPAPAKP-APTTPKETAPTTPKKTITTEBKLAFTTPKEKAPAPTPEELAP 451
DB 493 PS-----NPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 523
QY 452 TTPPEEPTTPPEEPAPTTPKAAAPNTPKEP-APTTPKEPAPTTPKEPAPTTPKETAFTTP 510
DB 524 SNPNP--PLNPNP-----SNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 564
QY 511 KGTAPTLKEPAPTTPKPAKELAPTTKEPTSTSDKPAFTTPGTAFTTPKEPAPTTP 570
DB 565 -----PSNPEPNPEE--PSNPKP-----SNPEEPINPELNPKEPSNPEESN 606
QY 571 PKPAPTTPKGTAPTTLKEPAPTTPKPAKELAPTTTGKPTTSDKPAFTTPKETAFTTP 630
DB 607 PKPPI-----NPEESNPKP-----INPEENPNLIIODEPIEPRNDSNVIFL 649

```

QY 631 TPKEP-----APTTKKRPAPTTETPPPTTSEVSTPTTKEPTTIHKSDESTPELSAE 684
 Db 650 LPIIFQKGNMIPNSLPEPNSDSEVEYPRPNDNGENSNNTMKSKNT---PNEPIPSGDN 706
 QY 685 PTKALENSPK-----EPGVPTTK 703
 Db 707 PYKGHEERIPKPHRSNDYVDYDNNVKNKNDPEIPNNE 745

RESULT 8
 PCT-US92-00018-2
 : Sequence 2, Application PC/TUS9200018
 : GENERAL INFORMATION:
 : APPLICANT: Hoffman, Stephen L.
 : APPLICANT: Charoenvit, Yupin
 : APPLICANT: Hedstrom, Richard
 : APPLICANT: Khushf, Richard
 : APPLICANT: Rogers IV, William O.
 : TITLE OF INVENTION: Protective malaria sporozoite surface protein
 : TITLE OF INVENTION: Immunogen and gene encoding
 : NUMBER OF SEQUENCES: 2
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: A. David Spevack
 : STREET: NMDC Building 1 T-12 National Naval
 : STREET: Medical Center
 : CITY: Bethesda
 : STATE: MD
 : COUNTRY: USA
 : ZIP: 20814-5044
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : OPERATING SYSTEM: IBM PC compatible
 : SOFTWARE: Patent Release #1.24
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: PCT/US92/00018
 : FILING DATE: 19920103
 : CLASSIFICATION: 424
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Spevack, Avram D.
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (301) 295-6759
 : TELEFAX: (301) 295-4033
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 826 amino acids
 : TYPE: AMINO ACIDS
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : PCT-US92-00018-2

Query Match 9.0%; Score 488.5; DB 5; Length 826;
 Best Local Similarity 26.4%; Pred. No. 3.5e-24;
 Matches 153; Conservative 54; Mismatches 223; Indels 149; Gaps 34;
 QY 168 TPTTISAPTTTKEPA-PTTTSAPTTTKEPA-PTTTKP-----APTTPKEP-APTTPKE 220
 Db 273 TCKVNDCCQIPIPIYIPKIKPEKPSNPEEPVNPDPNDPNNPNNPNNNNNNNNPNN 332
 QY 221 PAPTITKSAPTTKEP-APTTPKEP-APTTPKEP-APTTPKEP-APTTPKEPAPTTPKEPA 276
 Db 333 PNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNNDNSNPNPK 392
 QY 277 PTPKEPAPTAPK--KPATTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEAPT 334
 Db 393 RKNPKRRNNKPKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPNKPN 443
 QY 335 TTEPAPTITKSAPTTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 392
 Db 444 PAKPMP-----NPSNPNKPNP-----EPNPNESNPNESNPNESNPNESNPNESN 492
 QY 393 PAPTITKRAPAPAPKEP-APTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 451

Db 493 PS-----NPNESNPNESNPNNE--PSNPKK-----PSNPNNE--P 523
 QY 452 TTEEPPTTPEEPAPTTPKAAPNTPKPE-APTTPKEPAPTTPKEPAPTTPKEAPTTP 510
 Db 524 SNPN--PLNPNP-----SNPNESNPNESNPNNE--PSNPKK--PSNPNNE----- 564
 QY 511 KGTAPTTLKEPAPTTPKPAKELAPTTPKEPTSTSDKRAPPTTGTAPPTTKEPAPT 570
 Db 565 -----PSNPEEPNPEE--PSNKEP-----SNPEEPINPELNPKEPSNPEESN 606
 QY 571 PKEPAPTTPKGTAPTLTKEPAPTTPKPAKELAPTTPGTSTSDKRAPPTTKEAPT 630
 Db 607 PKEPT-----NPNESNPKPE-----INPDNPNPLIDDEPIEPRDSNVIP 649
 QY 631 TPKEP-----APTTKKRPAPTTETPPPTTSEVSTPTTKEPTTIHKSDESTPELSAE 684
 Db 650 LPIIFQKGNMIPNSLPEPNSDSEVEYPRPNDNGENSNNTMKSKNT---PNEPIPSGDN 706
 QY 685 PTKALENSPK-----EPGVPTTK 703
 Db 707 PYKGHEERIPKPHRSNDYVDYDNNVKNKNDPEIPNNE 745

RESULT 9
 5202236-13
 : Patent No. 5202236
 : APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
 : SUSAN L.; MCCANDLISH, RUSSELL; TENA, FILIPULA, DAVID
 : TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
 : PROTEIN
 : NUMBER OF SEQUENCES: 39
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/07/528,762
 : FILING DATE: 25-MAY-1990
 : APPLICATION NUMBER: 82,456
 : FILING DATE: 07-AUG-1987
 : APPLICATION NUMBER: 933,945
 : FILING DATE: 24-NOV-1986
 : APPLICATION NUMBER: 650,128
 : FILING DATE: 13-SEP-1984
 : SEQ ID NO:13:
 : LENGTH: 652
 : 5202236-13

Query Match 8.8%; Score 476.5; DB 6; Length 652;
 Best Local Similarity 29.7%; Pred. No. 1.6e-23;
 Matches 221; Conservative 71; Mismatches 284; Indels 169; Gaps 45;
 QY 51 PRTTAKPIINPRPSLPNSDTSKERSLTVNKEETVETETTTTKKOSTDKEKTTSAKE 110
 Db 11 PKMTPTPTTKKPKSPY-----TKSKPY-----KPKIT----- 40
 QY 111 TQSIEKTSKADLAPTSYLVAKPTPKAETTTGAPLTPPKKEPTTPKE-----PASTPKE 166
 Db 41 -----YPTTYK--AKPS-----YPTTYKPKTYPTTYKPLTYPTTYKPK 79
 QY 167 PTPITKSAPT-TKEPAPTTPKSAPTTPKAPAPTTPPTTKEPAPTTPKKEPAPTTPKEA--P 223
 Db 80 SYPTTYSKPTTKKTIYPTTYKAKPSYPTTYKPKKTYPTT-YKPKLTYPTTYKPKASYP 138
 QY 224 TTTTSAPTTTKEPAPTTPKAPAPT-TKEPAPT-----PKPEPTTTPKEPAPTTPKEAPT 278
 Db 139 PTVKPKPSYP--PSYKTKKTYPTTYKAKLTYPTTYKPKPSYKPKKTYPTTYKPKLT 196
 QY 279 TPKEPAPTAPKKA--PTTKEPA--PTTKEPAPTTPKEPSPTT-----KEPAPTTPK 329
 Db 197 YP-----PTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPT 252
 QY 330 SAPTTTKEPA-----PTTTSAPTTTPKEPSPTTTPKEA--PTTTPKEA-----PTTPK 376
 Db 253 AKPTTYKAKPKTYPTTYKAKPTYP-----PTYKAKPSYPTTYKAKPKTYKAKPKTYKAK 308


```

APPLICATION NUMBER: WO PCT/EP91/00835
FILING DATE: 23-OCT-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-08-479-537A-5

Query Match 7.8%; Score 424.5; DB 2; Length 1867;
Best Local Similarity 23.7%; Pred. No. 1,1e-19;
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

OY 75 TSLTVAKETTVETKETTINKOTSTDGKEKTTSAKETOSIEKTSKAD-LAPTSTVLAKPT 133
DB 16 TVLTV-----VTGSHASTPGEKETSATORSVSSSTKNAVSMTSSVLSHS 65
OY 134 P-KAETTTKGP--ALTTPKEP-----TPPTTKEPASTTPKEPTTTPKAP 176
DB 66 PGGSSSTTGGGCVTLAPATERASSAATWGDVTSVPVTRALGSTTPRAHDVT--SAP 122
OY 177 TTPKREPAPTTT-----KSAPTTPKEP-----APTTPKEPAPTTTPKEPAPT 217
DB 123 --DNRPAGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVT 180
OY 218 TKE--PAPTTP-----KSAPTTPKEPAPTTTPK-----KAPTTPKEPAPTTTPKEPTT 264

```

```

DB 181 APDXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVT 240
OY 265 PKE--PAPTTPKEP-----APTTPKEPAPTAK-----KPAPTTPKEPAPTTPKEPAPT 311
DB 241 APDXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVT 300
OY 312 TKEPSTTPKEPAPT--TKSAPTTPKEPAPT-----KSAPTTPKEPSTTPK-----E 360
DB 301 APDXRP--XPGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVT 359
OY 361 PAPTTPKEPAPTTPK-----KPAPTTPKEPAPTTPK-----EPAPTTPKRAPAPKEPAPT 412
DB 360 SAPDXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVT 419
OY 413 TPKETAP-----TPPKLTPTEKLAAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 468
DB 420 SAPDXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVT 471
OY 469 TPKA-----AAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 523
DB 472 APXAHGVTSAPODXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODXRPGS 530
OY 524 TTP-----KKPAPKELAPTTTKEPTSTTSKRAP--TTPKGTAPTTTPKEPAPTTP 571
DB 531 TAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODXR 585
OY 572 KPAPTTPKG-----TAPTTPKEPAPTTP-----KKPAPKELAPTTTKEPTSTTS 617
DB 586 PXPGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODX 644
OY 618 KPAP--TTPK-----ETAPTTPKEPAPTTP-----KKPAPTTPPTTPPTSEVSTP 661
DB 645 RXPGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODX 702
OY 662 TTTKEPTT-----IHKSPESTPELSAEPTRKALESPREBEVPTTPKPA-----707
DB 703 DXRPXGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODX 762
OY 708 -TKPMTTAAADK---TTERDLRTTPETTTAAPKMTKETATTTKEKTESKITATTTOVTS 763
DB 763 DXRPXGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODXRPGSTAPXAHGVTSAPODX 820
OY 764 TTDPTTPEKITTLLKTTTLLAKVTTTKTTITTEIMKKPRETAAPKRAATNSATTPKPO 823
DB 821 APDXRPGX-----STAPXAHGVT-----SAPDXRPGSTAPXAHGVTSAPODXRXP- 868
OY 824 KPTAPKKEPTSTKPKTPRVKPKTTPPKMTSTMBELNPTSRLEAMLOTTTRPNOT 883
DB 869 -GSTAPXAHGVTSAPODXRP--XPGSTAPXAHGVTSA--PODXRP-----PGST 911
OY 884 -PNSKLEVNPKSBDAGAGCEPNNMLLRPHVFMPEVTPMDYLLPRVNOGITINPLSD 942
DB 912 APXAHGVTSAPODXRPGSTAPXAH-----GVTSAPODXRPGSTAPXAHG 957
OY 943 ETNIGCKRPVDGLTT-LRNGTLVAFRGHYWMLSPSPSPSPARIRIEVWGIPSPID 997
DB 958 VTSAPDXRXPXGSTAPXAHGVTSA-----PDXRXPXGSTAPXAHGVTSAPODX 1003

RESULT 12
US-09-083-116-5
Sequence 5, Application US/09083116
GENERAL INFORMATION:
APPLICANT: CHAMBERN, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREVEENT, Maita
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATTHIS, L.L.P.

```

STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22131-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,116
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIOR APPLICATION DATA: WO PCT/FR91/00835
APPLICATION NUMBER: 23-OCT-1991
FILING DATE:
PRIOR APPLICATION DATA: US 08/039,320
APPLICATION NUMBER: 04-APR-1993
FILING DATE:
PRIOR APPLICATION DATA: US 08/403,576
APPLICATION NUMBER: 14-MAR-1995
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L. 35,030
REGISTRATION NUMBER: 017753-025
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note= "The amino acids spanning 128 to 1727 constitute a repeated region wherein the repeat is 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."
OTHER INFORMATION:
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."
OTHER INFORMATION:
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA or ACG; and Asn = AAT or AAC."
OTHER INFORMATION:
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."
OTHER INFORMATION:
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a 21 amino acid precursor sequence."
US-09-083-116-5

Query Match		7.8%;	Score 424.5;	DB 4;	Length 1867;
Best Local Similarity		23.7%;	Pred. No. 1.1e-19;		
Matches 255;		Conservative 69;	Mismatches 511;	Indels 241;	Gaps 53;
Qy	75	TSLVNKKETVETKETTITNKQSTDGKETTSAKETQSIKETSAND-LAPISKVLAKPT	133		
Db	16	TVLVV-----VSGSGHASTPGCEKETSATQSSVSSSTENKAVSMSTSVLSHS	65		
Qy	134	P-KAETTGGP--ALITPKPE-----PTTKEKPASTPKPEPTTIKSA	176		
Db	66	FGSGSTTGGQDVLALAPATEPAGSAAATWQDVTSVETPRALGSTITPPHMDT	122		
Qy	177	TTPKPEAPTTT-----KSAPTTPKPE-----APTTPKEAPTTKEPAPT	217		
Db	123	--DNKPAPGSTAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	180		
Qy	218	TKE--PAPTTT-----KSAPTTKEPAPTTPK-----KRAPTTPKPEAPTTKEPTPT	264		
Db	181	APDXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	240		
Qy	265	PKE--PAPTTKEP-----APTTPKEPAPTAPK-----KRAPTTPKPEAPTTKEPAPT	311		
Db	241	APDXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	300		
Qy	312	TKESPPTTKPEAPTT--TKSAPTTTKEPAPT-----TKSAPTTTKEPAPT	360		
Db	301	APDXRP--XPGSTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	359		
Qy	361	PAPTTTKEPAPTTPK-----KRAPTTPKPEAPTTTPK-----EPAPTTPKRAPTAKEPAPT	412		
Db	360	SAPDXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	419		
Qy	413	TPKETAP-----TTPKLTPTTPEKLAFTTPEKPAFTTPEELAPTTPEPTPTTPEEAPT	468		
Db	420	SAPDXRPXGTAAPXAHGVT-----APDXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	471		
Qy	469	TPKA-----AAPNTKEPAPTTPKPEAPTTTKEPAPTTPKPEAPTTTPKPEAP	523		
Db	472	APXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	530		
Qy	524	TPP-----KKPAPELAFTTTPKEPTSTSDKPAD--TPPKGTAFTTKEPAPTTP	571		
Db	531	TAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	585		
Qy	572	KEPAPTTPKGS--TAPTTTKEPAPTTP-----KKPAPELAFTTTPGPTSTSD	617		
Db	586	PXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	644		
Qy	618	KPAP--TPPK-----ETAPTTKEPAPTTP-----KRAPTTPETTPPTSEVTP	661		
Db	645	RPAHGSTAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	702		
Qy	662	TTTKEPTT-----IKSPDESTPELSAFTPKALENSKEGVTITTPAA-----	707		
Db	703	DXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	762		
Qy	708	--TKPEMTTAKDK--TTERDLRTTPTTAAKMTKEATTTTEKTESKITATTTQVTS	763		
Db	763	DXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	820		
Qy	764	TTTQDTTPEFKITLTKTTLAPKVTTKKTTTTEIMKPEETAKPKDRAATNSKATTPKPG	823		
Db	821	APDXRPXG-----STAPXAHGVT-----SAPDXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	868		
Qy	824	KPTAPKPTTTPKPT	883		
Db	869	GSTAAPXAHGVTSAPODXRP--XPGSTAAPXAHGVTSA--PODXRP-----PGST	911		
Qy	884	--PNSKLVENPKSEDAAGEETPHMLRPHEVMEVTPDMQDYLPRVNOGIIINPLUSD	942		
Db	912	APXAHGVTSAPODXRPXGTAAPXAH-----GVTSAPODXRPXGTAAPXAHGVTSAPODXRPXGTAAPXAHGVT	957		
Qy	943	ETINIGKRPVDCGTT--LNGTLVAFRGHYFWMLSPFSPSPARRTITVWGIPIPID	997		

QY 708 -TRKEMTTAKOK--TTERDLTTPETTTAAPKMTKETAATTEKTESKITATTTOVYS 763
DB 763 DAKRXPAGSTAPXAHGVTAPDXRXPAGST--APXAHGVTAPDXRXPAGSTAPXAHGVT 820
QY 764 TTTQDTPPKITTLKTTTLAPKVTTKTTTTEINMKPEETAKPKDRATNSKATTPKQ 823
DB 821 APDXRXPAGST--STAPXAHGVT-----SAPDXRXPAGSTAPXAHGVTAPDXRXPAGST 868
QY 824 KPKAKRKKPTSTKPKTMPVRKPTTPRKMTSTMPKELNPTSRJAEAMLQTTTRPNOT 883
DB 869 -GSTAPXAHGVTAPDXRXP--XGSTAPXAHGVTSA-PDXRXP-----PGST 911
QY 884 -PNSKLVENVPKSEDAAGETPHMLRPHVPEVTPDMYLPVPVNOGIIINPLSD 942
DB 912 APXAHGVTAPDXRXPAGSTAPXAH-----GVTAPDXRXPAGSTAPXAHGVTAPDXRXP 957
QY 943 ETNICKGKPVGLTT-LRNGTLVAFRGHTYWMLSFSPSPPARITEVWCIPSPID 997
DB 958 VTSAPDXRXPAGSTAPXAHGVTSA-----PDXRXPAGSTAPXAHGVTAPDXRXP 1003

RESULT 14
US-09-083-116-2
Sequence 2, Application US/09083116
Patent No. 6203795
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENEY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUVENT, Marie
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083.116
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ. ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning 128 to 1899 constitute a repeated region wherein the repeats varies from 1 to 40."
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION:
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
Xaa Xaa which is the codon for Pro or Ala wherein Pro = CC
CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or CGC."
OTHER INFORMATION:
OTHER INFORMATION:
OTHER INFORMATION:
OTHER INFORMATION:
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
which is the codon for Thr or Asn wherein Thr = ACT, ACC,
or ACG; and Asn = AAT or AAC."
OTHER INFORMATION:
OTHER INFORMATION:
OTHER INFORMATION:
OTHER INFORMATION:
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
which is the codon for Pro or Ala wherein Pro = CCT, CCC,
or CCG; and Ala = GCT, GCC, GCA, or CGC."
OTHER INFORMATION:
OTHER INFORMATION:
OTHER INFORMATION:
OTHER INFORMATION:
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-2

Query Match 7.8%; Score 424.5; DB 4; Length 2035;
Best Local Similarity 23.7%; Pred. No. 1.2e-19;
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 75 TSLTVNKETTVETKETTNTKORTSGKEKTSKAKETOSIEKTSAND-LAPTSTVLAKPT 133
DB 16 TVLTV-----VTSGHASTPGGEKETATATGRSSVPSSTKENVASMTSSVLSHS 65
QY 134 P-KAETTTKGP--ALTTTKEP-----TPPTTKEPASTTKEPTTTKSNAP 176
DB 66 PGSSSTTGQDVTALPATEPAGSAAWTGQDVTSPVTRPALGSTTPRAHVT--SAP 122
QY 177 TTREKPAPTT-----KSAPTTKEP-----APTTEKPAPTTKEPAPT 217
DB 123 --DKRPAPGSTAPXAHGVTAPDXRXPAGSTAPXAHGVTAPDXRXPAGSTAPXAHGVT 180
QY 218 TKE--PAPTTT-----KSAPTTKEPAPTTPK--KPAPTTPKEPAPTTPKPTT 264
DB 181 APDXRXPAGSTAPXAHGVTAPDXRXPAGSTAPXAHGVTAPDXRXPAGSTAPXAHGVT 240
QY 265 PKE--PAPTTKEP-----APTTPKEPAPTAPK--KPAPTTPKEPAPTTPKEPAPT 311
DB 241 APDXRXPAGSTAPXAHGVTAPDXRXPAGSTAPXAHGVTAPDXRXPAGSTAPXAHGVT 300
QY 312 TKEPSPTTPKEPAPT--TKSAPTTTKEPAPT--TKSAPTTTKEPSPTTTK--E 360
DB 301 APDXRXP--XGSTAPXAHGVTAPDXRXPAGSTAPXAHGVTAPDXRXPAGSTAPXAHGVT 359
QY 361 PAPTTPKEPAPTTPK--KPAPTTPKEPAPTTPK--EPAPTTPKAPATAPKEPAPT 412
DB 360 SAPXRPXPAGSTAPXAHGVTAPDXRXPAGSTAPXAHGVTAPDXRXPAGSTAPXAHGVT 419
QY 413 TPKEPAP-----TTPKKTLPPTPEKLPATTPPEKAPPTPEELAPPTPEPTTPEEPAPT 468
DB 420 SAPDXRXPAGSTAPXAHGVT--APDXRXPAGSTAPXAHGVTAPDXRXP--PGST 471
QY 469 TPKA-----AAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKGT--APTTLKEPAP 523

```

Db 472 APXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRP--XPGSTAPXAHGVTSAPDXRXPXGS 530
QY 524 TTP-----KKPAKELAPTTTKPTSTISDKRP--TTRGTPTRPKRAPPT 571
Db 531 TAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRXPXGS 585
QY 572 KEPAPTPPKG---TAPTTLKEPAPTPP-----KKPAKELAPTTTKPTSTIS 617
Db 586 PXPGSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDX 644
QY 618 KRAP--TTPK---ETAPTTKRAPPTP-----KKRAPTTPTPTTSEVSTP 661
Db 645 RXPGSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRXPXGS--TAPXAHGVTSAP 702
QY 662 TTKEPTT-----IHSPDESTPELSAERTPKALENSKPEPVPPTTKTAA----- 707
Db 703 DXRXPGSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAP 762
QY 708 -TKRPMTTAKOK--TTERDLRTTPTTAAAPKMKETATTEKTESKITATTTQVTS 763
Db 763 DXRXPGSTAPXAHGVTSAPDXRXPST--APXAHGVTSAPDXRXPSTAPXAHGVTS 820
QY 764 TTDDTTPFKTTTLTKTTLAKVTTTKTTTETIMKPEETAKPKDRATNSKATTPK 823
Db 821 APDXRXPXG-----STAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRXPX 868
QY 824 KPTAKPKPTSTKKRPMRVRKPTTTPPKMTSTMPELNPTSRILAMLOTTTRNOT 883
Db 869 -GSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRXPXGS-----PGST 911
QY 884 -PNSKLEVPNPKSDAGAGETPHMLLRPHVMEVTPOMDYLPRVNOGIIINPLSD 942
Db 912 APXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRXPXGS 957
QY 943 ETNINCKRPVGLTT--LRNGTLVAFRGHYFMMLSPSPSPAPARITVWMLIPSPID 997
Db 958 VTSAPDXRXPSTAPXAHGVTSAPDXRXPSTAPXAHGVTSAPDXRXPXGS 1003

```

RESULT 15
 US-08-276-967-2
 Sequence 2, Application US/08276967
 Patent No. 5851817
 GENERAL INFORMATION:
 APPLICANT: Hardy, Daniel M.
 TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of
 TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P. O. Box 4433
 CITY: Houston
 STATE: Texas
 COUNTRY: USA
 ZIP: 77210-4433
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/276,967
 FILING DATE: Submitted Herewith
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Kitchell, Barbara S.
 REGISTRATION NUMBER: 33,928
 REFERENCE/DOCKET NUMBER: UTS0418\KIT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 713-787-1400
 TELEFAX: 713-789-2679

```

; TELLEX: 79-0924
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2476 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-276-967-2

```

```

Query Match 7.7%; Score 419.5; DB 2; Length 2476;
Best Local Similarity 32.5%; Pred. No. 3,2e-19;
Matches 163; Conservative 43; Mismatches 177; Indels 119; Gaps 31;

```

```

QY 406 PKRAPPTPKETAPTTKPKLPTTPEKLAETPKRAPPTPEELAPTTPEEPPTPEEP 465
Db 313 PSETSVSTKRPVAPTE---KTPVPSIITYITPEKRVHMEKPIVHT--EKPTVPT-EKP 365
QY 466 APTTPKAAANTPKERA---PTTPKEPAPTTPKRAPPTPKETAPTTPKGAPTTLKBP 521
Db 366 TTPTEKSTVPT--KKPTVEKEPTLPPE-GPTVPAB-RPTTPREGPAVPPKG--PTVLTE- 418
QY 522 APTTPKPKAPKELAPTTTKEPTSTSDKRAPPTPKGTAP--TPKEPAPTTPKRAPPTPK 580
Db 419 -----WPTSHTEKSTVHTEKPLPTGKSTIPTEKRMVPTKRT-----TTP- 458
QY 581 GTPPTLKEPAPTTPKPKAPKELAP--TTKGPTST--SDKRAPPTPKETAPTTPKERA 636
Db 459 -TERTIIPAKKPIV---PIEKPMVPTERTIIPERTIIPTEKPTVTEKLIPT--EKPI 512
QY 637 PTPPKRAPPTTPPTP-----TISEVSTPT-----TKEPTTIHKSPESTPELSAERT 686
Db 513 VPEKPIVPEKHTIPEKLVLEKRTTPTERTIIPTEKPMVPEKPSVPT-EKPTVPT 571
QY 687 PKALENSKPEPVPPTKTPAATKPEMTTAKDKTTERDLRTTPETTTAPKMTKEATTT 746
Db 572 -----DEPTIPEKLIPT--ERTIIPKRTIITPTITRTIPTI-----RTTPT 613
QY 747 EKTTEKITATTTQVYSTTODTTPPKITTLTKTTLAPKVVTTTKTTTETIMKPEETA 806
Db 614 ERTT-----TPTIRTTPTERTT--IPTKKT-----VPEKTIIPT-----ERTI 652
QY 807 KPKDRATNSKATTPPKPKOKPTKAPKKPTSKKPKTMPRVAKKPTTTPPKMTSTMPELNPT 866
Db 653 AP-----TTPQ--SPTLVPTQPAVMPST-----SATVTPRTTIIASCP---PN 693
QY 867 SRIAEAMLOTTTRPNQTPNSKL 888
Db 694 AHFERCACPVSQ--SPTPNCCEL 714

```

Search completed: April 26, 2002, 16:17:34
 Job time: 245 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:18:59 ; Search time 82.85 seconds
(without alignments)
978.271 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1263

Sequence: 5698
1 VKDNKNRTRKKRTPKPPVY.....VAALSTAKYKNPESVFFK 1064

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR68:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1062.5	18.6	3020	2	A49332
2	950	16.7	1064	2	T18262
3	853	15.0	1489	2	T31108
4	802	14.1	1274	2	T16251
5	789.5	13.9	1367	1	S48478
6	784	13.8	1188	2	S49915
7	776	13.6	2187	2	T30826
8	672	11.8	1344	1	A35175
9	671	11.8	1151	2	T18535
10	659.5	11.6	1229	2	T25697
11	633	11.1	489	2	T11622
12	632.5	11.1	7962	2	T38346
13	632	11.1	761	2	C84672
14	629	11.0	3507	2	T34513
15	626.5	11.0	990	2	T51618
16	625	11.0	6642	2	T29757
17	622.5	10.9	839	2	T75518
18	622	10.9	3570	2	T45025
19	620.5	10.9	971	2	T19431
20	607.5	10.7	801	2	T29018
21	607.5	10.7	924	2	S27923
22	605	10.6	379	2	S50125
23	559.5	9.8	350	2	S22456
24	556.5	9.8	856	2	T16543
25	555.5	9.7	875	2	S23760
26	553	9.7	1630	2	A53577
27	551	9.7	2232	2	T44434
28	545	9.6	620	2	S06733
29	545	9.6	873	2	A47283

30	542	9.5	369	2	S20500	hydroxyproline-ric
31	538.5	9.5	416	2	JU0465	extensin precursor
32	533	9.4	1087	1	QPM5H	neurofilament trip
33	532	9.3	756	2	T27642	hypothetical prote
34	530.5	9.3	1162	2	JH0557	exo-alpha-sialidas
35	530	9.3	865	2	A47282	calcium-binding pr
36	522	9.2	328	2	J00985	hydroxyproline-ric
37	522	9.2	1459	2	T32271	hypothetical prote
38	518	9.1	813	2	S70795	vasa protein precu
39	518	9.1	866	2	T45462	membrane glycoprot
40	518	9.1	1072	1	A37221	neurofilament trip
41	512.5	9.0	867	2	T45463	membrane glycoprot
42	509.5	8.9	662	2	A45155	mucin FIM-C.1 - Af
43	508	8.9	1832	2	T31113	mucin-like glycopo
44	505	8.9	700	2	A54641	interpersed repea
45	504.5	8.9	606	2	A43427	neurofilament trip

ALIGNMENTS

RESULT 1
A49332
mucin 2 precursor, intestinal - human (fragments)
N:Alternate names: mucin SMUC-41
C:Species: Homo sapiens (man)
C>Date: 10-Mar-1993 #sequence-revision 12-Apr-1996 #text-change 05-Nov-1999
C/Accession: A49663; A45106; B45106; A43932; B35332; A61257; PQ0328; PQ0329
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A>Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of c
A:Reference number: A49663; MUID:94132002
A:Accession: A49663
A:Molecule type: mRNA
A:Residues: 1-639 <GU1>
A:Cross-references: GB:L21998
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J. Biol. Chem. 267, 21375-21383, 1992
A>Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up
A:Reference number: A45106; MUID:93016075
A:Accession: A45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 626-1895 <GU2>
A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396
A>Note: sequence extracted from NCBI backbone (NCBIP:116706)
A:Accession: B45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GU3>
A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398
A:Experimental source: colon
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J. Clin. Invest. 88, 1005-1013, 1991
A>Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
A:Reference number: A43932; MUID:91358717
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 143-1350, 'U', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
A>Note: sequence inconsistent with the nucleotide translation
A>Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A>Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evl
A:Reference number: A35332; MUID:89197956
A:Accession: B35332
A:Molecule type: mRNA
A:Residues: 1916-2193 <GU4>
A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
A:Experimental source: Intestine
R:Jenny, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

J. Clin. Invest. 87, 77-82, 1991
 A:Title: Human bronchus and intestine express the same mucin gene.
 A:Reference number: A61257; MUID:91086481
 A:Accession: A61257
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 17, 1925-1948, 'TTS', 1952-1954 <JAN>
 A:Experimental source: bronchus
 R.Xu, G.; Huan, L.; Khatri, I.; Saijan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner
 Biochem. Biophys. Res. Commun. 183, 821-828, 1992
 A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-
 A:Reference number: PQ0328; MUID:92198477
 A:Accession: PQ0328
 A:Molecule type: mRNA
 A:Residues: 2328-2468 <XUG>
 A:Cross-references: GB:M6523
 A:Experimental source: small intestine
 A:Accession: PQ0329
 A:Molecule type: protein
 A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
 C:Genetics:
 A:Gene: GDB:MUC2
 A:Cross-references: GDB:120203; OMIM:158370
 A:Map position: 11p15.5-11p15.5
 C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von
 C:Keywords: glycoprotein; intestine; tandem repeat
 F:2766-2834/Domain: von Willebrand factor type C repeat homology <WVC>

Query Match	18.6%	Score 1062.5	DB 2	Length 3020
Best Local Similarity	29.2%	Pred. No. 9,2e-39		
Matches 342	Conservative	74	Mismatches 442	Indels 315
				Gaps 39
QY 32	FKYTTTDDISTQTH---NKYSTSPKITTAKINDRPSILPPNSDSKSTSLVNKEITVEK 88			
DB 1240	WEICGNGVTEKHKHNCISITTRPSITLTFPTITLLPTTP-----TSFTTTTTTTPPS 1291			
QY 89	ET---TTTNK-----QSTDGKEKTTG--AKETOSIEKTSKAD----- 121			
DB 1292	STYLSSTPKICLCALMSWDINEDHPSGSDGDRPFDGVCAPEDIKCRKVDHLSTEOH 1351			
QY 122	-----LAPTSVLAKPPKAE 137			
DB 1352	GOKVOCDSVSGFLCKNEDQFNGFBGLCYDYKIRVNCQPMKCIITPSSPTTTPSPPT 1411			
QY 138	TTTKGPALTTPKPEPTPTTPKEPASTTPKEPTPTTIKSAP--TTKEPAPTITKSAPTTKE 136			
DB 1412	TTTTLPTTTPSPPTTTTTTPPTTPSPPTITTTTTPLPTTTPSPITSTTTTTTPPTTPPS 1471			
QY 197	PAPTTPKEPAPTTPKEPAPTITTKAPATTIKSAP--TTPKKEAPITPKKAPATTPKEPAPT 255			
DB 1472	P-PTTTPSPPTTTPSPPT--TTTTTPPTTTPSPPTTPTITTPASTITLLPPTTTPSPPTTT 1529			
QY 256	TPKKEPTPTTPKEPAPTTPKEPAPTTPKAPATTPKAPATTPKEPAPTTPKEPAPTTPKEP 315			
DB 1530	TTTTPTTTPSPPTTTPTTP--PTSTTLLPPTTTPSPPTTTTTTPPTTPSPPTTTPSP 1588			
QY 316	SPPTTPKEPAPTITKSAP--TTTKAPATTITKSAP--TTKEKESPTITTKAPATTTPKEPAPT 373			
DB 1589	PTTTTTPTTTPSPPTTTPPTTTTTTPPTTPSPPTTPTTPTTPPTTPPTTPPTTPSPPTTT 1648			
QY 374	PKKAPATTPKAPATTPTPKEPAPTTPKTKKAPATPKAPAKEPAPTTPKEPAPTTPKKAATPTEKL 433			
DB 1649	TTPTPTTTPSPPTTTPSPPTTTTTTTTTPPTTTPPTTTPSPPTTTP--SPPTT---TWTPPS- 1700			
QY 434	APTTPKEPAPTTPKEPAPTTPKEPAPTTPPTTTPKEPAPTTPKAAAPNTPKAPATTTPKEPAPT 493			
DB 1701	PTTTPSSPTTTPPTTPSSPTTP--SPPTTTPPTTPSPPTTTPPTTTPPTTTPSSPTTTP 1759			
QY 494	PKEPAPTTP-----KETAPTTP-----KG----- 512			
DB 1760	PLPSPITTPPTSPSTPTTTPPTCPVPLCMTWGTGLMSGAKNFHKPGSDTLTIGDVCPCGMAA 1819			

OY	513	-----	512
Db	1820	NISCRATMYDDVPIGOLGOTVWCDVSVLICKENEDOKGVIAPAFCLINEINYQCECV	1879
OY	513	TAPRTTLKEAPAPTPPKKPAKRELAAPT--L-KEPTSTSDPK-----APT--TRPGAPAPTPKE	565
Db	1880	TOPTTM---TTTTENTPTPIPTTTTTVTPPTPTPTSTOSPAGLOAPTPPTSTTTTTVTPPT	1936
OY	566	PAPPTKEPAPAPTPKPGAPPTTLKEAPATTPPKKPAKRELAAPTTTNGPST--TSDKPAPTP	624
Db	1937	PRPTGOTPT--TPTTTTTVTPPTPTGOTPTVLTITTTMTPTPTSTKSTVTP	1995
OY	625	KEIAPAPPKAPAPTPPKP-----APPTPTPTPTSEVSTPTTK-----EPTTLHK	672
Db	1996	ITTTTTVAPPTPTGOTPTPMIPITSTTTTTVTPPTPTGOTPTPTSTADIALTLTNSP	2055
OY	673	SPDESTPELSAEPPTKALENSPKREPOVPTTKTPAATKAPMTTADKTTERDLRTTPEPT	732
Db	2056	PPESPTPTSTASSTPSPTTESTTTLSTLPPAIEKTSIAPRTSTLAPTTSGCHTLSPST	2115
OY	733	TAAAPMTKEATTTTTTEKTTESKITATTTVOVSTTTTTODTTPFKITT--LKTTTLAPKVTTT	789
Db	2116	TTSPTGPTTRGTTT--GSSAPPTSTVQTUUTTSAMPTPTPTSTPSIIIRITGLRPBSSV	2173
OY	790	KKTITTTIEMKPEB-----TAKPKDAFNTSKATPT	820
Db	2174	LKCVLINDTYAPAGEEYVNGTYGDTCYVNCISLCTLEFYNNMSCBSTSPPTPTSK--STP	2232
OY	821	KPKOPTKAPKKPTSTKPKKPTMPVRKPKTPTPTPKMTSTMPBELNPTSRILAEMLOTTTPR	880
Db	2233	TTSKASTSPSKPTPTOTKRPCEPDPPR-----QENETMMLCDOCFM--ATCKY	2278
OY	881	NOTPNSKLEVENP-----KSDAGC-----AEGTPTMML-----	910
Db	2279	NMTVELVAKVECEPPMPTCSNGLOPVRVEDPDGCMHMECDYCTGWDG--PHYVTFDGLY	2337
OY	911	-----RPVFMPEVTPPTMDOLPRVPNNOGIIN	937
Db	2338	YSYOGNCYVLEVEITSPSVD-----NKGVIID	2364

```

RESULT      2
T18262
S:Layer protein - Clostridium thermocellum
C:Species: Clostridium thermocellum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18262
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e:
A:Reference number: Z18847; MUID:93209931
A:Accession: T18262
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1,1664 <FNU>
A:Cross-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA4784.L.1

Query Match          16.7%: Score 950; DB 2; Length 1664;
Best Local Similarity 31.0%: Pred. No. 3, 8e-34;
Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;

Oy    128 VLAKPTP-KETTTKGALTTPEKPPPTTKKEPASTTPKEPYTTTKSAPTTKKEAPATT 186
       | : ||| ||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| :
Db    758 VVIQPAPIKAASDPIPTDTPSDPTPS-----DEPTPS--DEPTPSDEPTPSD 804

Oy    187 TKSAPTTKKEAPTTTKKEAPATPKKPA----PTTKKEAPATTKSAPTTKKEAPATTK 242
       : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    805 EPRTSETPPEEPITPDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTSPSETPPEEPI 864

Oy    243 KPAPTTPKEAPATPKKEPPTTKKEAPATTKKEAPAT-TKKEAPATPKKPAPTTPKEAP 301
       ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

```

Db 865 SDEPTPSDEPTPS--DEPTPS--DEPTPS-SDEPTPSDEPTPSDEPTPSDEPTPS 919
 QY 302 TPKKPA--PTTKKPSPT--TPKKEAPPTTKS-APTTKKPAAPT--TTKSAPTTPKPS 354
 Db 920 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 979
 QY 355 ----PTTKKPAAPT--TPKKEAPPTPKKPAAPTTPKKEAPPTTPKKEAPPTTPKKEAP 409
 Db 980 PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1039
 QY 410 APT--TPKKEAPPTPKKPAAPTTPKKEAPPT--TEBELAPTPPEPTPT-TP 462
 Db 1040 TPSPDEPTPS--TPKEEPTPTDTPSDEPTPSDEPTPSDEPTPSD-EPPTPSDEPTPS 1094
 QY 463 EEPAPPTPKKAAPTTPKKEAPPTTPKKEA--PTTKKPAAPT--TPKKEAPPTPKKPAAPT 519
 Db 1095 EEPPTPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1154
 QY 520 EEPAPPTPKKPAAPTTPKKEAPPT--STTSKPAAPTTPKGAAPTTPKKEAPPTTPKKEAPPT 578
 Db 1155 EEPPTPS--DEPTPSD-EPPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS--DEPTPS- 1208
 QY 579 PKGAAPTTPKKEAPPT--PKKPAAPTTPKKEAPPTTPKKEAPPT--TPKKEAPPTTPKKE 634
 Db 1209 ----DEPTPSDEPTPSDEPTPS--PTDTPSDEPTPSDEPTPSD-EPPTPSDE 1253
 QY 635 PAAPTTPKPAAPT--TPETPTPTSEVPTTPKKEPTTIHKSPDETPBELSAEPTPKALENS 693
 Db 1254 PTPS--DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS--PSDEPTPS--SDEPTPSDEPTPS 1305
 QY 694 PKKPAAPTTPKPAAPT--KPKMTTAAOKTTERDLKPTTETTTAAAPKMTKETAATTPKETE 751
 Db 1306 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD--EPPTPSDEPTPSDEPTPSDEPTPS 1361
 QY 752 SKTATATTTTQVTSSTTQDTPPEKITTTLKTTTTLAKVTTKKITTITTTIMKKPEETAKPKR 811
 Db 1362 DEPTPSDEPTPSDEPTPSDEPTPS--TTPTPTPTPT--TTPTPTPTPT--TTPTPTPTPT 1396
 QY 812 ATMSKAT--TPKOKKPTKAP--KKPTSTKKKPTKAPKPKKPTTPPKM-- 856
 Db 1397 SGSGSGSGGGGGGGGCTVPTSPPTPTSKPTSPAP--TEIEPTPSDVPGLIGEENH 1453
 QY 857 ----TSMPELNTPSRIAEAMLQTTTRNPTPKSKLVENPKSEDAAGAGETPHN- 908
 Db 1454 YLRGYDGSFRPRNITRAEAIVF-----AKLL--GADESYGASQSPYS 1498
 QY 909 LKRPH-----VFMP-----VTPDMDYLPVPMOGI----- 934
 Db 1499 LADTHMAAMAKFATSGLFKGYPGDTFRPDQNTIRAEATVVLHFLTVKGOEIMSKLA 1558
 QY 935 ----IIPMISDETNICNG--KPYDGLTTL-----RNGTLVAFRGHYFMMLSPPSPS 981
 Db 1559 TDISNPKFD--CVGHMAOEFEKLTSLGYSIGYPDGT-----FKPQN 1600
 QY 982 PARITEVWGISPIDVTFRNCGEKGTPFEKD--SOVYRFTNDIKD 1026
 Db 1601 YIKRSSTV--ALINRALEKRLGAPKLPDVNESTIAP--GDMID 1642

RESULT 3
 T31108
 cyst germination specific acidic repeat protein precursor - Phytophthora infestans
 C:Species: Phytophthora infestans (potato late blight agent)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T31108
 R:Goenhardt, B.
 Submitted to the EMBL Data Library, April 1998
 A:Reference number: z20986
 A:Accession: T31108
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1489 <GOE>
 A:Cross-references: EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AAC72308.1

C:Genetics:
 A:Gene: car90

Query Match 15.0%; Score 853; DB 2; Length 1489;
 Best local similarity 32.3%; Pred. No. 5.3e-30;
 Matches 355; Conservative 46; Mismatches 475; Indels 222; Gaps 49;

QY 9 TTKKPPKPPVVDAGSGLDNDGKVTPTDSTT-----QHNKYSTSKITTAAPINR 62
 Db 338 TPYAPTEKEDYDEETTYVTEESTYAPTKSETNAPTRMHYAIKPKOEVIMVAPTEET 397
 QY 63 PSLP-----PNSDT-----SKESLVYKKTIVYETETTT--TNKOTSTD 100
 Db 398 TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 457
 QY 101 GKEKTTSAKETOSIEKTSADLAPTSKVLAKPTPAEETTKGAPALJTTPKEPTTPKKA 160
 Db 458 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 514
 QY 161 STTPKEEPTTTIKSAPTPKKEAPPT--TKSAPTPKKEAPPT--TPKKEAPPTPKKPAAPT 217
 Db 515 EETTPYEETET--TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETTPYEET 573
 QY 218 TKEPAPTTKKSAPT-----TPKKEAPPTPKKPAAPTTPKKEAPPTTPKKEAPTTPKE 267
 Db 574 TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETTPYEETTYAPTEETTYA 632
 QY 268 P-----AP--TPKKEAPPTPKKPAAPT--APTAPKPAAPTTPKKEAPPTTP 304
 Db 633 PTEETTYAPTEETTYAPTEETTYAPAEETTPYEETTYAPTEETTYAPTEETTYAPTEET 692
 QY 305 KKPAPT--TKKPSPTTKKPAAPT--APT--TKSAPT-----TPKKEAPPTPKKPAAPT 339
 Db 693 TTYAPTEETTYAPAEETTPYEETTYAPTEETTYAPTEETTYAPAEETTYAPTEETTYA 752
 QY 340 PT--TKKSAPT--TPKPSPTTKKPAAPT--APTAPKPAAPTTPKKEAPPTTPKKEAPPT 388
 Db 753 PTEATTYAPTEETTYAPAEETTPYEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 812
 QY 389 TPKKEAPPTTKKPAAPTTPKKEAPPTTPKKEAPPTTPKKEAPPTTPKKEAPPTTPKKEAP 448
 Db 813 TYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 870
 QY 449 LAPT--TPKEEPTTPKKEAPPT--APT-----TPKAAAPNTPKKEAPPTPKKPAAPT 489
 Db 871 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 930
 QY 490 ----APTTPKKEAPPTPKKPAAPTTPKKEAPPT--TLKKEAPPTPKKPAAPT--ELAPT--TT 539
 Db 931 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 990
 QY 540 KEPTSTSDKAPPTPKKPAAPTTPKKEAPPTPKKPAAPT--APTTPKGAAPT----- 585
 Db 991 YAPTEETTYAPTEETTYAPTEETTYAPAEETTPYEETTYAPTEETTYAPTEETTYAST 1050
 QY 586 ----TLKKEAPPTPKKPAAPT--APKE--LAPT--TTKGSPTSDKPAAPTTP 624
 Db 1051 EETTYAPTEETTYAPAEETTPYEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1110
 QY 625 KETAPTPKKEAPPTPKKPAAPTTPKKEAPPTTPKKEAPPTTPKKEAPPTTPKKEAPPTTPK 682
 Db 1111 YAPAEETTPYEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 1169
 QY 683 AEPKPA--LENSPKKEG-----VPTTKPAATKEMMTTAKKOTTEDLNLT--ETT 732
 Db 1170 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1229
 QY 733 TAAFKMTKETATTEKTESKITATTTQVSTTQDPT-----PKKIT--LKT 779
 Db 1230 TYAP-----TEETTYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1284
 QY 780 TTLAPKVTTKKITTITTTIMKKPEETAKPKDRAAT--NSKATTPKP--OKPTKAPKPKPT-- 833

```

Db 1285 TTYAPLEATYAPTEETPYAPTEETPYGTYAPTEETTYAPTEETPYAPTEETPYE 1344
OY 834 -----STKRP-KTTPRRVKKPKTPPKMTSTMPPLNTPSKRAEMLQTTTRP--N 881
Db 1345 PAESTSTVSTKPCMTTEETDEPTDEPTDE--PSDEPTDEPTDEPTDEPTPCDN 1402
OY 882 QPNSTKLVNPKSEDAG 899
Db 1403 QGTINGIVENKRYNNAG 1420

```

RESULT 4

```

116251
hypothetical protein F35A5.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C:Accession: T16251
R:Leimbach, D.
submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid F35A5.
A:Reference number: Z18485
A:Accession: T16251
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1274 <LEI>
A:Cross-references: EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AA852641.1; GSPDB:GN00
A:Experimental source: strain Bristol N2; clone F35A5
C:Genetics:
A:Gene: CESP.F35A5.1
A:Map position: X
A:Introns: 1272/2

```

Query Match 14.1%; Score 802; DB 2; Length 1274;

Best Local Similarity 28.0%; Pred. No. 7.2e-28;

Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;

```

OY 6 KNRTRK-KP-----TP-----KPPVVDAGSGLDNGD-----FKYTTDTSTQHNKSTKS 50
Db 274 KNPKKMKPPEWEDTPVEEVKEPPVPEKKAPVLKKKDPAPAKADSPSKAPKVEPS 333
OY 51 PKITTAPRI-NRPSLPNSTSSTSETSJLVNKETVETKETTNNKOTSTDGKETSATK 109
Db 334 SPVVPPEPVKPKPVKKRPVEVDEPRAEVKKPSAPEKKTPVLKKEPESSSTPSSDPS 393
OY 110 ETOSIEKTSANDLAPTSKVLAKPTPKAE-----TTTKGPA-----LTTT 148
Db 394 PKKAAPAVKPRDSSPKKATPLQADPKAOEVPPTVKNPKKKYKPPWEVDEDPVEEVKOP 453
OY 149 KEPPTPTP-----KEPASTTPEP-----TP-TTISAPTTPEKPAPTTKSAPTPKKEPAP 199
Db 454 EAPAKKPTVLKKEPAADTKAPATSKPTPEPEKKDPVKPRDSSPKVAAKPDSAQAP 512
OY 200 TTTKEPA-----PTTKEPAPTTTKEPAPT-----TTKSAPTTP 233
Db 513 TPVKVPVKMKRPWEDETPADVDVSKPTDAKKTSIAKKDAPAKESLSKPADTKAPAPK 572
OY 234 KEP-----APTTPKKPAPTTPE-----PAPTTPEKPTTTTKEPAPTKEPAPTTPK-EP 283
Db 573 RDPSPKKAAPTAPKEKTPVLAKKEPAGPADSKTEKEPEKSPRDPSPKAAVPAKVPKTEV 632
OY 284 APTAKKKAP-----TTPKEPAPTTPKEPAPTTPKEP-----SPTTPEKPAPTTKSAPT 333
Db 633 AAAAAYKKBEPSKPKDTPAKKAEPSVVP-PTPVKNVKKMKPPWEDDAPAKVSLPE 691
OY 334 TTKPEAPTTPKSAPTTP-KEPSPTTTPKEPAPTTPK-EPAPTTPKKAP--TTPKEAPVT 388
Db 692 PEKK-TPVLAKKAPTKPDSAAADVSGSPSKDPLAKKAPVKPRDPSKAVPITKAPAK 750
OY 389 TTKPEAPTTPKKAPPA-PA-PKEPAPTTPKETATTPKTKLTPPTP-----EKLAPTPEKPA 442
Db 751 T--EVPAPAVVKKPEVAKSRDPSKKAK-AEENSP--VVPPTPVKNVKKMKPPWEDDA 805

```

```

OY 443 PTTPEELAPTTPEED-----TPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPT--PKE 496
Db 806 PAEPVNVPEPEPKKTPVLAKKTPVKPRDPSKRAVAPASTKTDAPVSVKKEPEVSKPE 865
OY 447 PAPTPKTEAPPTTPGTAATTLKEPAPTTPPKKAPAKELAPT-TTKEPTSTSDKPAPTTP 555
Db 866 PSPKKAEPNSPVVP-----PTPVKNPVKKK-KPPEWDEDEPEVKKKSE--PEKKTPLA 918
OY 556 KGTAPTTKEPAPTTPKPEPAPTTPGTAATTLKEPAPT-----TTKKAP-----KEL 603
Db 919 K-KEPEKPD-APKVAARPRDPSKKAPE--KEPAKVAARPRDLSPKALIPANTQEA 974
OY 604 APTTKGPTIS-----TSDKP-----APTTPKET-----APTTPKEPAPTTPKKAPPT 646
Db 975 PTPVKNPVKKMKPPWEDDDEPAEVSAPKEPEKTPVLAKKAPAKPRP-----SPKKAAPV 1031
OY 647 TPEPTPTTSEVSTPTTKEPTTIKSP-----DESTPELSA-EP--TPKALENSPKPEG 698
Db 1032 AAK-FDPKRIPEV-PTTPVKNPVKKMKPPWEDDDESESVASPEPEKKTPLAKKAPTKPA 1089
OY 699 V-----PTTKTPAAT-----KPEMTTAKDKTERDLRTTPTT--TAAPK 737
Db 1090 TKPDEAAADPVSGPTSKDPLSKRAPEKRPPTTDPKDDKIKPSPAKKPEKAPAPAK 1149
OY 738 MKKETATTEKTESKITATTTQVSTTQDTTPPKITTLKTTTLAPKVTYTKKTIITTE 797
Db 1150 KMKPWWDDPDDEPAEFTYVAPSKKPDIEDPDLG-----GPKTKDPK----- 1193
OY 798 INNKPEETAKPKDRATNSKATTPPKQPKTPAKPKTSTIKKPKTPMRVKKPK----- 848
Db 1194 -LNKKAPAKKPTKE-----PKREVSKPEPKPTKEPEPK-AAPKKMKPPWEDDPDDEPE 1243
OY 849 ---TTPTPPKMTSTMP 861
Db 1244 ADEFTMPAPKKPDIEDP 1259

```

RESULT 5

```

S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR01
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C:Accession: S48478; A26877; B26877; S27281; JG6123
R:Rowley, K.
submitted to the EMBL Data Library, October 1994
A:Reference number: S48478
A:Accession: S48478
A:Molecule type: DNA
A:Residues: 1-1367 <ROW>
A:Cross-references: GB:Z47047; EMBL:Z38061; NID:9603997; PID:g763364; GSPDB:GN00009;
J. Bacteriol. 169, 2142-2149, 1987
R.Yamashita, I.; Nakamura, M.; Fukui, S.
A:Title: Gene fusion is a possible mechanism underlying the evolution of STR1.
A:Reference number: A91831; M01D:87194600
A:Accession: A26877
A:Molecule type: DNA
A:Residues: 1-242 <YAM>
A:Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
A:Accession: B26877
A:Molecule type: DNA
A:Residues: 762-1331 <YAZ>
A:Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A:Title: Similar short elements in the 5' regions of the STR2 and SGA genes from Sacc
A:Reference number: S27281; M01D:89031230
A:Accession: S27281
A:Molecule type: DNA
A:Residues: 1-31 <PAA>
A:Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R:Leimbach, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.

```


Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A:Title: Muc1, a mucin-like protein that is regulated by Msi10, is critical for pseudohy
A:Reference number: J06123; MID:9633237
A:Accession: J06123
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1367 <LAN>
A:Cross-references: GB:030626; MID:g1304386; PIDN:AC49609.1; PID:g1304387
C:Gene: SGD:MUC1, STRA2; MAL5; DEX2; SGD:S0001458
A:Cross-references: MIPS:YIR019c; SGD:S0001458
A:Map position: 9R
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C:Keywords: glycosylase; hydrolase; polysaccharide degradation; transmembrane protein
F:5-21/Domain: transmembrane #status predicted <TM>
F:1350-1366/Domain: transmembrane #status predicted <TM>

Query Match 13.9%; Score 789.5; DB 1; Length 1367;
Best Local Similarity 28.4%; Pred. No. 2.7e-27;
Matches 301; Conservative 104; Mismatches 483; Indels 171; Gaps 47;

```
QY 35 TTPDSTTQHNKYSTSPKITTAKP-----RSLPSPNSDTSKESLVNKEITYE 86
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 236 STESSSTTSSSTSSSTSSSTTAPAPPTTSCPKKPTPTTSCPKKPTPHHDTPC 295
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 87 TKEPTTNNKQSTDGKEKTTSAKETOSIEKTSADLAPTSKVLAKPTPKAETT--TKGA 144
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 296 TKKTTTTSK-TCT--KTTTTPPTPS--STTSSSAPV-----PTSSSTTSSSAPV 344
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 145 LTTPKE-----PTTPKEPASTTPKEPTPTTKSAP--TTPKEAPPTTKSAPTP 194
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 345 TSSSTSSSAPVPTSSSTSSSAPVSTTSSSAPVSTTSSSAPVPTSSSAPVPTSSSSTE 404
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 195 KEPAPTTKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTP 254
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 405 SSSAPVPTSS--TTESSSAPVTS--STTESSSAPVTS--STTESSSAPVTSSTES 453
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 255 TTPKEPTP--TTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTP 307
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 454 SSAAPVPTSSSTTSSSAPV--SSTTESSSAP--VPTSSSTTSSSAPVSTTSSS 509
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 308 APTTTPKEPTTPKEPAATTPKSAPTTTKEPA--TTTKAPPTTPKEPTTTPKEPA 363
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 510 AAVPTPS--STTESSSAPAPVPTSSSTTSSSAPVSTTSSSAPVPTSSSTSSSP 568
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 364 ----TTPKEPAAT--TPKAPATTPKEPAATTPKEPAATTPKAPATTPKEPA 418
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 569 VTSSTTSSSAPVPTSSSTTSSSAPVPT--PSSSTTESSSAPAPTPSSSTTESSA 624
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 419 PTTPKLTPTTPEKLAPTTPKEAPTTPEELAPT--TPEEPTTPTEEPAP-----TPKA 472
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 625 PVT---SSTTESSSAP--VPTPSSTTSSSAPVPTSSSTTSSSAPVPTSSSTSS 679
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 473 AAPNTPKKEAPPTTPKEPAATTPKEPAATTPKETAAT--TPKGAATTLKEPAATTPK 531
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 680 SAPV--SSTTESSSAPV--SSTTESSSAPVPTSSSTTSSSAPVPTSSSTTE 731
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 532 KLAATTTKEPTSTSDKAPATTPKGAATTPKEPA--TTPKEPAAT--TPKGAATTL 587
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 732 SSSAPVPTPS--SSTTESSSAPVSTTSSSAPVPTSSSTTSSSAPVPTSSSTSS 790
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 588 KEPAATTPKAPKLAATTLTGPTSTSDKAPATTPKETAATTPKEPA--TTPK 642
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 791 SAPVPTSSSTTSSSAPVPTSSSNTSSAPSTPTSSSTSSSAPVPTSSSTSS 850
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 643 PAPPTTPPTTSEVSTPT-----TKEPTTIHKSPEDESTPELSAETPKALENSPEP 697
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 851 SAPVSSSTTSSSAPVPTSSSNTSSAPST--PSSSTTESFSTGT--TPTSSSKATP 906
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 698 GVPPTTKAARKPEMTTAKDKTERDKRTPEPTTAAPKATKETAATTEKTESKTTAT 757
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 907 GSOTETSVSTTEETTVPTKTTTSTPTTTTITTVOSTGTGAGETSGCSKPTVTTT 966
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```

```
QY 758 --TTOVSTTQDTPPEKITTTLTKTTLAPKVT--TKKTI--EIMNKPEETAKPKOR 811
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 967 VPTTTTSVTSSTTITTTVCSGTGNSAGETTSKSPRTITTTTPCSTSPSETA----- 1021
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 812 ATNSKATTTKPKPKP-----TKAPKPTSTPKPTMPVRKPKTTPPR 854
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 1022 ---SESTTSPPTPVTVTVSTVTVTEVSTSTKPGCEITTVTFVKNIPTTYLTITAPR- 1077
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 855 KMTSPPELNPSTRIAEAMLQTTTRPNQNPNSKLVENKSPDAGAGETPHMLLRHV 914
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 1078 -SVTVTNTPTTT--ITTVCSGT-----NSAGETTSKSPKTVTTTVCST 1122
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 915 EPEVTPMDMVPVPPVNOGIIINPMLSDETNICKPKVDTGLTTLNGLTVAFRGHYFWM 974
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 1123 GTGEYTTTEATLVTTAVTTVTTSTSTGNTSA--GKTTGTGTTKSVPT-----YVITL 1175
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 975 SPSPSPPARRTTEWGPSPIDVFTRC---NCEKGT 1009
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 1176 APSAPVPTATN-----AVPTTIT--TECSATATNAGET 1207
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```

RESULT 6
S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C:Accession: S49915
R:Rutinshtein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Rex genes: pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Accession: S49915
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: EMBL:Z34465; MID:g600117; PIDN:CA84230.1; PID:g600118

Query Match 13.8%; Score 784; DB 2; Length 1188;
Best Local Similarity 29.9%; Pred. No. 4e-27;
Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

```
QY 13 PTPKPPVVDGASGLDNGDFKVTTPDSTTQHNKVNSTPKTTAKPTNPRSL----PPN 68
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 460 PTPHSPRAD-----DIVPPTPVGKSPATSPSPQVQPPAASPTPSLKKLSPQ 510
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 69 SDTSKETSLVKNKETVEKETTNNKQSTDGKEKTTSAKETOSIEKTSADLAPTSKY 128
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 511 APVG-----SPPPVKTTSPAPIG-----SPSPPPVSVV 541
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 129 -----LAKPTPKATTTKGPALTPKKEPTTPTPKEPASTTPKKEPTTYSAPTPKEPA 183
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 542 SPPEPVVSPPEPAVGVSPPEPKSPPEPAVASPPPVKSP--PPPLVASPPPVKSP 599
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 184 PTTKSAPTPP--KEPAATTTKEPAATTPKEPAATTTKEPAATTTKSAPTTT--KEPAPT 240
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 600 PPAVASPPPVKSPPTPTVASPPPAVASPPPKKSPPTPTVSSPPPEKSPPEPP 659
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 241 PKKAPPTTPKEPAATTT-----PKE--PPTTPKEPAATTPKEPAATTPKEPAATPK 291
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 660 PAKSTPEPEVPTPTPVSVKSSPPEKSLPPTLIPSPPEKPTLPSTPEK--PSSPEK 718
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 292 APTTPKEPAATTPKEPAATTTKESPTTPKEPAATTTKSAPTTTKEPAATTTKSAPTTTRK 351
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 719 SP--PKPPVSPPTP-----KSSPPPAVSSPPTPVSSPALAPVSSPVSASP-- 768
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 352 EPEPTTTPKEPAATTPKEPAATTPKAPATTPKKEPAATTPKEPAATTTTKKAPAPKKEPA 411
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 769 PPAFLSSPPAPQVKKSPPVQVSSP--PPAPKSSPLAPVSSPPQVEKTSPPAPLSSPP 827
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 412 TTPKEPAATTTTP-----KKLTPTPEKLAATTPPEKPA--TTPPELAATTPPEEPT 458
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```

```

Db      838  IAPKSPRHVVVSSPPVVKSSPPAPVSSPPLTKRASPRAHVSSPPVVKSPSP - -PA 885
OY      459  PTTSEAPATTPKAARNTPKE - PAPTTKEBAPTTPKBAPTTKEIATTPKGTAPTT 517
      111  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      886  PTTVPSP -PSEKSSPPPTPVSLPPPIVKSPPAMVSSP -PMTKSSPPPVVSSSPT 943
OY      518  LKE - - - - -PTTPPK - - -PAPKELATTTKE - - -PTSTSDKAPTTPKGTAPTP 563
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      944  VKSSPPAPVSSPPATPKSSPPAPVNLPPPEVKSPPPTVSSPPA - -FKSSPPAP 1000
OY      564  -KEAPATTKEBAPTTPKGTAPTTLKEBAPTTPKKAPKELATTTKGTSTTSDKAPAT 622
Db      1001  MSSPPPEVKSPPAPVSSPPPVKSSPPAPVSSP - - -PPVKSPPAPVSSPPPV 1057
OY      623  -PPKEATTPKEBAPTTTKKAPATTPEPPPTTSEVPTTKKEPTTIHKSDSTPEL 681
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1058  KSPPPAPATSSPPPVKSSPPAPVS - -SPPPVKSSPPAPVSSPPPIKSSPPAPVS 1115
OY      682  SAEPTPKALENSKEGVPVTTKTPAATKPEMTTATADKTERLRTTPET 731
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1116  SPPPA - - - - -VKPSSLP -PPAPVSSPPVVTAPAPKKEOSLPPPAES 1158

```

RESULT 7
 T30826
 nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
 N:Alternate names: alpha-NAC protein
 C:Species: Mus musculus (house mouse)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
 C:Accession: T30826
 R:Yotov, W.V.; St-Arnaud, R.
 Genes Dev. 10, 1763-1772, 1996
 A>Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle
 A:Reference number: Z20889; MUID:96312450
 A:Accession: T30826
 A>Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Molecule 1: 1-2187 <YOM>
 A:Cross-References: EMBL:U48363; NID:q1666688; PID:q1666689; PIDN:AA818732.1
 C:Genetics:
 A:Gene: Naca
 A:Map position: 10
 A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
 A:Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activ
 C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match	13.6%;	Score 776;	DB 2;	Length 2187;
Best Local Similarity	27.7%;	Pred. No. 1.6e-26;		
Matches 288;	Conservative 122;	Mismatches 389;	Indels 240;	Gaps 49;

QY	34	VTTDSTQHNKXVTSPTKTTKPLNPRSLUPENDSTKE-----	TELVTNKEPTT	84	
Db	806	VOSKRVUDIMSDVTPISPKTSATAV-PK-----	DTSATLSLKSVAIVSLSPKAPV	857	
QY	85	VEIKETTTNKQSTDGKEKTTSAKETOIEKTSANDLAPTSKVLAKPPKAEITTRKPA		144	
Db	858	APSNEATIVPTPIETSLKNALAAATPKETLATSIPVTSIS-----	PQTKPSVSLKQAPV	913	
QY	145	LTT-----PKK-----	PTPTTKPEASTT-----	PKKEPTTT	171
Db	914	MTSKKATEIASKDVSPSOPKREVLLOHVPTSPKSPVSDLSGALNSPPKGP-PAT		972	
QY	172	IKSATPTTKE-PATTTKSAPTTK-----	EPAPTTTKEPATTTKEPATTTKEP	221	
Db	973	IAETTTKSPKPAASKTPTATPSPGCVAVLEILPCKSKNAPKMAKESASISSKR		1032	
QY	222	APT-T-TKSAPT-----	TKPEAPTTPKK-PATTTKEPATTP	257	
Db	1033	AKPTAVSEIKSVGTAVPLEISLPLKETSASATPGKSSASRSRSPATAGKE-----	TPP	1086	
QY	258	KEPTTTTKEPATTTKEPATTPKEPATTPKAPATTT--PKKPAP-----	TTPKKEPATTT	311	

```

Db      1090 GGVAVAPPEISLPKPTPQNTPNESIQAASSQKSPSTSVPKETPPGCGVTAAMPLEIPSA 1149
Qy      312 TKESPSTPK-----EPAPTTTKSAPTTTKBPAPTTT-KSAPTTPK-----EBSPTT 357
Db      1150 QKAPRTAVPKQILPPEDAVNTILAGSPILSPKASKTAAPKEAPAPSGVIAVSGEISIP 1209
Qy      358 TKEPAPTTPKPEAPITTPKKPAPT--PRE-----PAPTTKPEAP-----T 396
Db      1210 KTSKTAAPKKNASATLPPKRSPPKTAAPKEPTAISSEGVTAVPSLISSPITPASKGVPT 1269
Qy      397 TTKKPAPAPKEPAPATTPPKETAPTTPKLPTPEEKLAPTTPEK-PAPTTPEELAPTTPE 455
Db      1270 LTPKAPAPALAE-SPASQKVKVPTAAPBETSTP-----SPQKIPVAGKEKASAPPS 1322
Qy      456 EPTPTT-----PEEPAPTTPKAAANTKEKAPITTPKE-PAPTTPKKEPAPITPKETA 506
Db      1323 KKTPTAVPKETISASEGVTAVPLEIPSPPKRAPKTAAPKETPAPS--PEGATTAAPQIP 1380
Qy      507 PTTPGTAPTTLKEBAPTT-----KKPAKELAPTTKEPTS-- 544
Db      1381 PEPKRGSKKAGKE-TPTTTPSEGVTAAPLEIPSSKTSKMSAPKEPLVPPSSKKLSQ 1439
Qy      545 ----TTSOKPAPTTPKGTAAPTTPKEPAPITTPKE-PAPTTPKGTAAPTLKEBAPTTPKPA 599
Db      1440 VGPKESTLEGAAVAVLEIPSSHKAAPKTVDPQVLTSPK-DATITLAE-SPSSPK-A 1496
Qy      600 PKELAPTTTKGPTSTSDSOKPAPTTPKETAPTTPKEPAPITTPKKAPITPETP-----PPT 654
Db      1497 PKTAPPSR-VTYVPEKPA-TPOKASGTAASKVPAVEIQEVAVSRETPVPAVPPV 1554
Qy      655 TSEVSTPTTTK-----EPTTHKSPDESPTPLSAEPTPKALENSPK-PGVPTTKPAA 707
Db      1555 KPPSSHKRSKTELEKAPATLPPSPKSPKIPSSKKAAPT--SAPKEEPASPSTK-PVT 1611
Qy      708 TKPEMTTTPKDKTTERDLPTTPPETTTAAPKMTKETATITTEKTESKIAATTTQVYSTTQ 767
Db      1612 T-SLAQTAPPSIQKAPSTTIFIKENLAAPV----LVPSSKSPAPAPARASLASPTAAP 1665
Qy      768 DTTPEKITTLLKTTTLAAKVTTTKITTTTTTINKKEPEETAKPPDRATINSKATTPKQKPTK 827
Db      1666 QTPAKKATITPSCKAATAETPELETSTAPSLGAKETSE-----TSSKVLMSSP----- 1716
Qy      828 APKPKSTTKPKPMVRKRPKTTPTPRKMTSMPELNPSTRIAEAMLOTTPRPNQTPNSK 887
Db      1717 -PKKASSSKRASLP-----ATLLPSLKEASVLS-----PTATSSGK 1752
Qy      888 LVEVNPKSEDAAGABEETP 906
Db      1753 DSHISPPS-DACSTGTTP 1770

```

RESULT 8
A:35175
mucin 1 precursor, repetitive splice form A [validated] - human
N:Alternate names: Breast carcinoma-associated DF3 antigen; core protein KP39; episial
ncretic mucin; polymorphic epithelial mucin (PEM)
N:Contains: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor
N:Species: Homo sapiens (man)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Jun-2000
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Jun-2000
C:Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; P40066; S10218;
J. R. Littenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hlikens, J.
J. Biol. Chem. 265, 5573-5578, 1990
A>Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene
A:Reference number: A35175; MUID:90202794
A:Accession: A35175
A:Molecule type: mRNA
A:Residues: 1-952,103-1344 <LIg1>
A:Cross-references: GB:M2728; GB:005286; NID:q182121; PIDN:AAA5804.1; PID:q182124;
A:Experimental source: splice form A
A>Note: Genbank entries HUMEP1A1 and HUMEP1A2 present only the amino and carboxyl
A:Accession: B35175
A:Molecule type: mRNA
A:Residues: 1-19,29-952,103-1344 <LIg2>

A:Cross-references: GB:M32739; GB:J05288; NID:9182126; PIDN:AAA5806.1; PID:g182129; GB:
A:Experimental source: SPLICE form B
A:Note: Genbank entry HUMPEP18 and HUMPEP18B present only the amino- and carboxyl-ter
R:Gendler, S.J.; Lancaster, C.A.; Taylor-Pedemonte, J.; Dunlap, T.; Peat, N.; Burchell
J. Biol. Chem. 265, 15286-15293, 1990
A:Title: Molecular cloning and expression of human tumor-associated polymorphic epithel
A:Reference number: A35886; MUID:90368715
A:Accession: A35886
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19,29-992,1033-1344 <GEN>
A:Cross-references: GB:J05581; NID:9188669; PIDN:AAA59876.1; PID:g188870
A:Note: Genbank entry HUMWUCAB includes one copy of the tandemly repeated sequence
J. Biol. Chem. 265, 15294-15299, 1990
A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
A:Reference number: A35887; MUID:90368716
A:Accession: A35887
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
A:Cross-references: GB:J05582; NID:9189598; PIDN:AAA60019.1; PID:g189599
A:Note: Genbank entry HUMPANUW contains four fewer copies of the tandemly repeated sequ
R:Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.
Eur. J. Biochem. 189, 463-473, 1990
A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera
A:Reference number: S10571; MUID:90276413
A:Accession: S10572
A:Molecule type: mRNA
A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <WR2
A:Cross-references: EMBL:X52229; NID:937053; PIDN:CAA6478.1; PID:g37054
R:Wreschner, D.H.
submitted to the EMBL Data Library, March 1990
A:Reference number: S40293
A:Accession: S40293
A:Molecule type: mRNA
A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <WR2
A:Cross-references: EMBL:X52229; NID:937053; PIDN:CAA6478.1; PID:g37054
R:Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.
Eur. J. Biochem. 189, 463-473, 1990
A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
A:Reference number: A36735; MUID:90088473
A:Accession: A36735
A:Molecule type: mRNA
A:Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>
A:Cross-references: EMBL:M31823; NID:9181542; PIDN:AAA35757.1; PID:g181543
R:Masuzawa, Y.; Myauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.
J. Biochem. 112, 609-615, 1992
A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu
A:Reference number: JX0235; MUID:93123189
A:Accession: PX0066
A:Molecule type: mRNA
A:Residues: 998-1011,'E',1014-1017,1018-1032,'T',1034-1037,1038-1057 <MAS>
A:Experimental source: gastric carcinoma cell
R:Zirhan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
FEBS Lett. 356, 130-136, 1994
A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine
A:Reference number: S51026; MUID:95080414
A:Accession: S51026
A:Contents: annotation
A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region at
C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 C
partial repeats. The repeat shown is defined by SmaI nuclease sites.
C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
C:Genetics:
A:Gene: GDB:MUC1; PUM
A:Cross-references: GDB:120705; OMIM:158340
A:Map position: 1q21-1q23
A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
C:Superfamily: polymorphic epithelial mucin
C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphis
F:1-1344/Product: mucin I precursor, splice form A #status predicted <PREA>
F:1-62/Region: mucin I amino-terminal non-repetitive

F:1-23/Domain: signal sequence #link PREA #status predicted <SIGA>
F:1-19,29-332/Domain: signal sequence #link PREB #status predicted <SIGB>
F:1-19,29-1344/Product: mucin I precursor, splice form B #status predicted <PREB>
F:1-19,29-212,1033-1344/Product: mucin I precursor, epithelial tumor antigen splice f
F:138-1017/Region: 20-residue repeats (GSPAPAHGVTSAPDRAP)
F:1143-1344/Region: mucin I carboxyl-terminal non-repetitive
F:1245-1272/Domain: transmembrane #status predicted <TRM>
F:1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predic
F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 11.8%; Score 672; DB 1: Length 1344;
Best Local Similarity 27.8%; Pred. No. 3,1e-22;
Matches 310; Conservative 78; Mismatches 472; Indels 256; Gaps 56;

QY 77 LVNKKETVEKETTNTNK--OTSDGKEKTTSAKETOSIEKTSKD-LAPTSKVLAKPT 133
DB 15 LVNLTATTAPKPAIVVTSQGHASSVTPGGEKETSAQRSSVSTKKNVSMVSSVLSHS 74
QY 134 P-KAETTKGP--ALTTEKP-----TPPTKEPASTTP-----KE 166
DB 75 PEGSSSTTQGDVTLAPATEPAGSAAATWGODVTSVPYTRPALGSTTPPAHDVTSAPDNK 134
QY 167 PPTPT-----IKSAPPTKEPAPTTP-----KSAPTPKEPAPTTPKEP-----A 206
DB 135 PAPGSTAPPAHGVTSAPDT--RPAGGSTAPPAHGVTSAPDT--RPAGGSTAPPAHGVTS 190
QY PPTTKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTP-----KKPAPTTPKEP----- 252
DB 191 PDT--RPAGGSTAPPAHGVTSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGSTAPPAHGV 247
QY 253 --APPTKEPAPTTP-----KEPAPTTPKEP-----APPTKEPAPTTPK----- 289
DB 248 TSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGSTAPPAHGV 307
QY 290 KPAPTTPKEPAPTTPK-----EPAPTTPKEPAPTTPKEPAPTTP--TSAPPTTPKEPAPT 342
DB 308 TSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGSTAPPAHGV 367
QY 343 TKSAPTTPKESPPTTPK-----APPTKEPAPTTPK-----KPAPTTPKEPAPTTPK- 391
DB 368 T-SAPDT--RPAGGSTAPPAHGVTSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGSTAPPA 424
QY 392 --EPAPTTPKPAPTAPKEPAPTTPKETAAPTTPKILPTTPREKLAPPTTPKEPAPTTP- 447
DB 425 HGVTSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGSTAPPA 484
QY 448 --ELAPPTPEEPPTTP-----EEBPAP--TTPKA-----AABPTTPKEPAPTTPK- 487
DB 485 HGVTSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGSTAPPA 544
QY 488 --EPAPTTPKEPAPTTPKETAAPTTPKETAAPTTPKETAAPTTP-----KKPAPKEL 534
DB 545 HGVTSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGST 600
QY 535 APPTTKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 568
DB 601 AP-PAHGVTSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGSTAPPAHGVTSAPDTTP 655
QY 589 EPAPTTP-----KKPAPKELAPPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 632
DB 656 APGSTAPPAHGVTSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGSTAPPAHGVTSAPDT 714
QY 633 KEPAPTTP-----KKPAPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTP 675
DB 715 PAPGSTAPPAHGVTSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGSTAPPAHGVTSAPD 772
QY 676 ESPPELSAEPPTPALENSPEKGVPTTPKTPAA-----TKPEMTTJAKDK---TTERD 724
DB 773 TRPAGGSTAPPAHGVTSAPDTTPAPGSTAPPAHGVTSAPDTTPAPGSTAPPAHGVTSAPD 832
QY 725 LKTTPTPT-----TAAKMTKETATTEKTESKITATTTQVTSSTTQDPTTP--FKIT 775

Db 833 TRPAGSTAPPAHGYTSAP----DTRPAGSTAPPAGVTSAPDTRPAGSTAPPAGHVT 888
QY 776 TLTETTLAPKVTITTKITTTT--EIMNKPDEETAPKDRATNSKATTPKPKMTAKPKKPT 833
Db 889 SAPDTRPAGSTAPPAGHVTSAPODTRPAGSTAPPAGHVTSAPODTRPAP--GSTAPPAHG 946
QY 834 STKKKPTMPRVKPKTTPPTPKMTSTMPELND---TSRIAPAMLOT--TTRP---NOTPN 885
Db 947 VTSAPDTRP---APGSTAPPAGHVTSAPODTRPAGSTAPPAGHVTSAPODTRPAGSTAPP 1003
QY 886 SKLVEKNPSEDAAGAEGETPHMLRPHVEMBEVTPMDYLPVRVNOGITIINPMLSDFTN 945
Db 1004 AHGYTSAPDTRPAGSTAPPAPH-----GVTSAPDNKRALGSTA---PPVHNVTIS 1049
QY 946 ICNGKPVDTLLRNGTLVAFRGHYFWMLSPPSPPS 981
Db 1050 ASGSASGSASTLVHNGTSARATTTAPKSTPPSPIS 1085

RESULT 9
T18535
high molecular mass nuclear antigen - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18535
R:Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick
A:Reference number: Z18955; NUID:9803440
A:Accession: T18535
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1151 <SH1>
A:Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BA24137.1

Query Match 11.8%; Score 671; DB 2; Length 1151;
Best Local Similarity 25.9%; Pred. No. 3e-22;
Matches 284; Conservative 100; Mismatches 419; Indels 292; Gaps 55;

QY 14 TPKP-----PYVDAGSGIDNGDFKVTTPDSTTOHNVKSTS--PKITTAKPIPRRS 64
Db 113 TTPPSQGPAGTTPPSQGAAGAPKGDGTAGPSGTGSGADGKPAADVPKATTA--ATARP- 170
QY 65 LPPKSDTSKSTSLTVNKETIVERKETTTNK--QTSIDGKEKTTSAKETOSIEKTSAKDL 122
Db 171 -----ASAASPTVPKATDEATAVTAASQASAPKAADDAVAATA--SOSAKAIY-EV 219
QY 123 APTSKVLAKPTPKAETTTKGPALTTPKEPP-TPPKEPASTTPKEP---TPTTIKSAPT 177
Db 220 KPAAAVAKAKAVTAATAAAKPAATAEKAPAVTSPTIPCSAEAKPLTASPTASKA--T 277
QY 178 TPKEPAPT-----TKSAPTTPKEPAPTTK----- 203
Db 278 AEAPKPVATASLMAKTATAEKAPAPSPSVKATDTKAVTATAPKAGPDVKAIVAVCAEA 337
QY 204 EPATTPKEPAPTTPKESAPT-----PKEP-----PTTKKPAPTPK 250
Db 338 KPAPPPPOOLPKAAAAAATPGTLKATATAPPHCSPPRANHVTIVTPPNVPRAAAATVP- 396
QY 251 EPAPTTPKEPPTT-----PKEPAPTTPKEPAPTTPKEPAPPAKPPADTTTPKEPAPTTPKE 306
Db 397 -TAGAVPKASTGTTPAAPAOQVPV---KAAPVTTPPSPOQAVPRAATAAA---APVTPQO 448
QY 307 P---APTTPKEPSTTPKEPAPTTPKESAPTTPKEPAP-----TTTSAPTTPKEPSTTPKE 360
Db 449 PVTKAATTTNTATPPQOLPKAATTTTATPVTPQOLPKAGTDAAPPAVPAASDGAAT 508
QY 361 P---APTTPKEPAPTTPKPAPTTPKEPAP-----TTTPKEPAPT--TTTKKPAAT 406
Db 509 PGVPRAAADDPKPPPTPOSVSATETPKQRAAPPSNEATPAVPSPSPLKSLDPTIP 568
QY 407 K-----EPAPTTPK--ETAPTTPKKLPTPEKLAPTTPKEPAPTTPPELA----- 450

Db 569 KPVPLMALITQPVTAQWOTLAATKPSIYPKASPK-ALMTPEPPPGGLRALAAKLLG 627
QY 451 -PTTP-----EETPTTP---EEPAPTTP-----KAAPPTPK----- 480
Db 628 LPSSVASAMHAKVTPPLPASPVPMAASPSLGDPAARVALATNAATNAPAKPEAAAGNG 687
QY 481 -PAPTTPKEPAPTTPKEPAPTTPKEPAPT-----TPKGTAPT----- 516
Db 688 TLMAFMGAANTOMAPIGAGAQAOTAPMGAANTHVSPPMGAGATOMSTGGAANTHMSPIGA 747
QY 517 -----TLKEPAPTTPKPPAPKELAPTTTKEP-----TSTTSKPAPTTP 555
Db 748 GGATOMSPMGAANTOMSPMGATTTOMSPMGAATTTQSPMGAATQYATISAGNTMQVSP 807
QY 556 KG--TAPTTPKEPAPTTPKEPAPTTPKGTATTTKEP--APTTPKPAPELATTTTGP 611
Db 808 MGAATPPQTPSVGAATTP--QSPM---GAATTLMSPGCAATTPQ---PSPMCAVTTQPP 859
QY 612 -----TSTSDKP-APTTPKET---APTTPKEP---APTTPKEP-APTTPPETTP----- 652
Db 860 PMAATNTTQPPMAASTTQTPMGAATTTQSPMGAATTTQSPMGAATPPQAPPTVAGSPT 919
QY 653 PTTSEVSTPTTTKEPTTIHKSPPDESTPEL--SAEPTKALENSKREQVPTTKTPATKP 710
Db 920 PPPPIPPSPATQTSQPMKSPPPDPKAPSAQAOTSPAIVANAAPGV-TAVSPA---P 975
QY 711 EMTTAKKKTTERDRTPEPTTTAPAKMT-KEIATTTKTESKITTATTOVSTTTTQDT 769
Db 976 IGVTAESPADGARSPPGTLAATDGPKASPAATADVTAAID--VTAATAVPA-----EA 1029
QY 770 TPKITTLTKTTTLAPKVTITTKTITTTTEIMNKPETAPKDRATNSKATTPKPKAP 829
Db 1030 AP-----TKAKRSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSESNPASP- 1077
QY 830 KKPSTKKKPTMPRVKPKKTTTPPKMTSTMPELNPTSRIMEAMLOTTRNQPNKSLV 889
Db 1078 -----PAVGDSQOQONTPGAQSVPP-----VTEAAVQ----- 1104
QY 890 EYNPKSEDAGAGE 904
Db 1105 EAAAAAAGAGARE 1119

RESULT 10
T25697
hypothetical protein F16F9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T25697
R:Fulton, B.
submitted to the EMBL Data Library, August 1996
A:Description: The sequence of C. elegans cosmid F16F9.
A:Reference number: Z20071
A:Accession: T25697
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1229 <FULL>
A:Cross-references: EMBL:067956; PIDN:AA07691.1; GSPDB:GN00028; CESP:F16F9.2
A:Experimental source: strain Bristol N2; clone F16F9
C:Genetics:
A:Gene: CESP:F16F9.2
A:Map position: X
A:Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 11.6%; Score 659.5; DB 2; Length 1229;
Best Local Similarity 29.0%; Pred. No. 1e-21;
Matches 307; Conservative 78; Mismatches 339; Indels 333; Gaps 52;

QY 33 KVTT-PDTSSTVO---HNKVSIPKTTTA-KPINRPSLPPNSDT-----SKETSLTVN 80
Db 119 KVTTSIDASTTNAPJTQKDSSTPELITGIVINSKSESVTDMSTRFSTTSLSPTELLNS 178

```

QY 81 KETVETKETTNTKQSTDKERTTSKAKENQSTKTSKADLAPTSKVLAKPPTKAEETT 140
D 179 PETLVSDDSSSTST-EQTSPODNTTELASPMENINTTEATNTSVESVSGLAS---EDETIV 234
QY 141 KGPLMLTPEKPEPTTPKPEAPSTTKEPTPTTKSAPPTKPEAPPTTKSAPPTKPEAPT 200
D 235 TALEST-----TTLVAESTTTEETPTT-----ASTTKKSTT-----KAPA 272
QY 201 TTKPEAPPTKPEAPPTTKPEAPPTTKSAPPTKPEAPPTP-----KKRAPPTKPEAPPT 256
D 273 TTEETPTPTTEE--VTTEATSTTTSSSESTEK---PTTBLDNKTAGPATG---PET 324
QY 257 PKEP-TPTTKPEAPPTKPEAPPTTKPEAPPTKAPKRAPPTTKPEAPPTKPEAPPTTKE- 314
D 325 THFVGTGTTPN--FDIATEPTFVAKSSEDKMTLSKATATEGTQOTEVT-DGPEKETIKV 361
QY 315 ----PSPTTKPEAPPTTKSAP-----TTTKPEAPT-----TTKSAPPTKPEPSPTTK 359
D 382 SIEPTTTLVETSTSTTSKESDGFHTTLKLVTTADSDSTESATVTVKPTNEETTK 441
QY 360 E---PAPT-----EPAPTPKRAP-----TTPK 383
D 442 SHVVPKPTKGGYKVTPLKELSDDEPLEIT-KAPHKGLLEKTYHPVLSDNARISEAK 500
QY 384 E-----PAPTPEKPEAP-----TTTKRAPTA--PKPEAPPT---PKETAP 419
D 501 ENDYNLHDYNYHREAKEPTTTESSSTTEETVTEEPANGNPTENPTTEQPTSTAE 560
QY 420 TPEKKLPTTPEKLA---PTTPEKAPPTPEELAPPTPEPTPTTPEEPAPPTPKAAPN 476
D 561 STTALPPTTEQVITTEETPTAKSTATO---KPTTQSVST---EKSTTKKA---S 610
QY 477 TPEKPEAPPTKPEAPPTTKPEAPPTTKGAPPTPKGAPPTTKPEAPPTPKRAPKELAP 536
D 611 TTEB--PTTDEPTT---ESSTGKATPBELSTSEETTELKITE-----GS 657
QY 537 TTKKEPSTSDKRAP---TTPKGAAPTTPKPEAPPTKPEAPPTPKGAPPTTKELAP 592
D 658 TTEETPTTALFAEASTGITTDEETSTSTPEITSTKE--IVMESAITQSVSVES 715
QY 593 TTP-----KKRAPKELAPTTKGP-----TS 613
D 716 STPRLPERKKAIVNKFKNHLEVLAKKKLKEKSTSTSDSSSTTTVAENIDEVT 775
QY 614 TTSOKAPPTPKETAPPTKPEAPPTPKRAPPTPE--TPPTTSE--VSTPTTKKEPTI 670
D 776 TEKEKVVQTPPTITTEKSTQOETTTTTEKTKSTTEKPTTSATTEITTTSEPT- 834
QY 671 HKSPDSTBELSAEPTPKALENSPKPEGVPTTKTPAATPEMTTAKDKTTERDLTPE 730
D 835 ----TEST-----TVDTSATTEESSTAETTTSAE---TSE 865
QY 731 TTTA-----APMTKETTTTEKTE 751
D 866 TTTSESAFATGSPENTALOSSOKSENESSAEKPGARDFVKKKHTVTKPAETTSA 925
QY 752 SKTATATTO-VIST---TQODTPEKITTLKTTTAPKVT---TTKKTITTEINKKEE 804
D 926 VASTTTPPTTTEKSTLTETTPTEATTLNVENTGPAPVTCGAPVDETITNTELLSK--- 982
QY 805 TTKPKDRATNSKATTPKPO-----KTPKAKKCTSKKPKPTMRRVKKP 847
D 983 -----INNTQISQKPPIDISKTOALSGLIGSTFKAMAPTIT----- 1022
QY 848 KTTPTPKKMTSTMPELN-----PTSRIAEA 872
D 1023 HTTDAAPVATATEASLNDGSKKIIDEAQPTEIRRA 1059

```

RESULT 11
T11622
extensin class 1 precursor - cowpea

C:Species: Vigna unguiculata (cowpea)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000
C:Accession: T11622; S54135
R:Arsemljevic-Maksimovic, I.; Broughton, W.J.; Krause, A.
Mol. Plant Microbe Interact. 10, 95-101, 1997
A:Title: Rhizobia modulate root-hair-specific expression of extensin genes.
A:Reference number: 217301; MUID:97155574
A:Accession: T11622
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-489 <ARS>
A:Cross-references: EMBL:X91836; NID:g1015936; PIDN:CA62943.1; PID:g1015937
A:Experimental source: sub.species Red calcoana
R:Arsemljevic-Maksimovic, I.; Broughton, W.J.; Krause, A.
Submitted to the EMBL Data Library, April 1995
A:Description: A class of root-hair specific extensins involved in rhizobium/legume
A:Reference number: S54155
A:Accession: S54155
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 326-489 <AR2>
A:Cross-references: EMBL:X86030; NID:g791149; PID:g791150
C:Genetics:
A:Gene: Ext26g
C:Superfamily: hydroxyproline-rich glycoprotein
C:Keywords: glycoprotein; hydroxyproline
F.1-23/Domain: signal sequence #status predicted <SIG>
F.24-489/Product: extensin class 1 #status predicted <MAT>

Query Match 11.1%; Score 633; DB 2: Length 489;
Best Local Similarity 31.2%; Pred. No. 5.7e-21;
Matches 149; Conservative 33; Mismatches 252; Indels 44; Gaps 7;

```

QY 179 PKEPAPPTTKSAPPTKPEAPPTTKPEAPPTTKPEAPPTTKSAPPTTKSAPPTKPEAP 238
D 39 PKQTPPYVYNAAPPYVYKSP-----PSPSP-----PPPYVAKYPPYVYKSP 83
QY 239 TTPKRAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPT 295
D 84 PSPSPPYVYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 141
QY 296 PKEPAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKSAPPTTKPEAPPTTKSAPPTTK 355
D 142 YKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 187
QY 356 TTKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPT 415
D 188 YVYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 247
QY 416 ETAPPTPKKLPTTPEKLAPTTPEKAPPTTPEELAPTTPEETPTTPEEPAPPTPKAAP 475
D 248 PPPPYVYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 307
QY 476 NTPKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKGAPPTTKKAPAPTTKAPKELA 535
D 308 PSPSPPYVYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 366
QY 536 PTTKEPSTSDKRAPPTTKGAPPTTKPEAPPTTKPEAPPTTKGAPPTTKKAPAPTTK 595
D 367 ----KSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 422
QY 596 KKPAPKELAPTTTKGPTSTSDKRAPPTTKGAPPTTKPEAPPTTKKAPAPTTPEPTPP 653
D 423 SP-----PPVYVYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 475

```

RESULT 12
I38346
elastic tftin - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C:Accession: I38346

A:Reference number: 151618; MUID:96019267
A:Accession: 151618
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-990 <CAI>
A:Cross-references: EMBL:X88927; NID:9895920; P1DN:CAA61368.1; PID:9895921
C:Genetics:
A:Gene: xNoppl80
C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

Query Match 11.0%; Score 626.5; DB 2; Length 990;
Best Local Similarity 26.2%; Pred. No. 2.2e-20;
Matches 253; Conservative 132; Mismatches 393; Indels 187; Gaps 45;

QY 3 DNKKNTKKKPPPPVVDAGSLDNGDFKTTPTSTQHNKYSTSPKITTAKPINR 62
DB 59 DAKRRRPANGLPKRRSAKSSSSSEDEPPAKKRAQ----PAGKKPVKAVQPK 114
QY 63 PSLPNSDTSKETSIVNKETIVE--TKETTTNK---QSTDGKETTSAKETQSIK 116
DB 115 KAKSSSEDSDESD---SEETKKPPAKRPAOTPKVAAYKTPQKKAKSSS-ESSSED 170
QY 117 TSAKDLPATSKVLAKPPKAETTTKGPALTTPEKEPTTPKE----PASTPEPTPTI 172
DB 171 EAKSKKOPVIKV----PPKQAVVAGIASNNGTADSSSESDSPPAKKTATKTPPT- 225
QY 173 KSAPTTPKEPAPTTTSAPTTKEPAPTTTKEPAPTTTKEPAPTT--TKEPAPTT-----T 226
DB 226 --KPAITAAPKPAKTTAKKSSSSREDSDEQOKTAKSKPKPDVYSAVPPITSVKKT 283
QY 227 KSAPTTPKEPAPT-----TPKKPAPTTKEPAPTT---KEPTPTTPKEPAPTT 272
DB 284 LSGPGTAKESSSDSSDSEEQPAKAKIYPAKAAASAPKLAETSTDESDSS 343
QY 273 KEPAPTTPKEPAPTPAKKPAKTPKEPAPTT---KEPAP-----TP 311
DB 344 EDEKSSSVKLGVKAAPK--APAAP--DAKSTPVAAKKSAAPAKKASSSDSSSNEET 400
QY 312 TKESPPTTPKEPAPTTTSSAPTTTKEPAPTTTSSAPTTTKEPAPTTTKEPAPTTKEPA- 370
DB 401 TKPAKATTPKASATPTSKPTNKG---ATPSTKPAKPGTPTSTAKKSSSDSS 457
QY 371 ---PTTPKPAPTTPKEPAPT--TPKEPAPTTTTPKAPAPKERA--PTTPKEIA----- 418
DB 458 SDEETTTKPAKTPAKSAATPTSKTP---TNSKATPTSKTPAKPGTPTSAKKDSS 514
QY 419 -----PTTPKLTPTPEKLAAPTPPEKAPATTPPELAPTPPEEPTPTTPEEPAPTP-KA 472
DB 515 SDSSDSSDEKTPA--KRAKTPAKPA-----AKTTPAKPA-----AKTTPAKP 558
QY 473 AAPNTPKEPAPT---TPKEPAPTTTPKEPAPTTTKEPAPTTT- KGIAPTTTKEPAPTTPK 527
DB 559 AAKSTPGQVPTKKSSSDSSDSSSEDEKSSAKPAVKTPGKATS-----KPVVAS 611
QY 528 KPAPKEIAPTTTKEPTSTTSKPAPTPKGTAPTTPKEPAPTT--PKEPAPTTPKGTAAPT 585
DB 612 KPVPAK-----KASSSDSDSSEETTKTKPLKLSPAVKTLPPKAESSSDSSSDSD 665
QY 586 TLKEBAPTTTPKPAKELAPTTTGP-----TSTTSKAPAPTTTKEPAPT--TPKEP 635
DB 666 SEKK---TKPAKPPAKSATPVNTKAPAKNKAASKASCSDDSSSEEGSKOPTGKSPA 722
QY 636 APPTPK--PAPTPPTPTPTTSEVSTPTTKE--PTTIHKSPEDESTPELSAAPTPEKALENS 693
DB 723 ATAPPKKPNVAVNKDPSSSSSDSDGDEKQPKQAAAAADVQGAANAAPTPPKAAS 782
QY 694 PKPEGVPTTKAATKPEMTTAKDTERDLRTTETTTAAPTKEATATTEKTESK 753
DB 783 SSE---DSSSDSDVSKAKKNTAVSKSPV---TTPKAVPAKAKSSSESDSEDEKOG 835
QY 754 ITATTTQVSTTODTTPFKITTLKTTLAPKVTTK--KITTTVEIMNKPEETA-KPKD 810

DB 836 KNSTTKIANST-----PKAAAAECSESSSSSEDEGKANGTSGKRR 877
QY 811 RATNS---KATTPKPOKPTAPKPKPISTKKPKTIPRKRKKTPTP-RKNTSTMPELNPT 866
DB 878 ESTGNACEAVTPE-----NKKLKAKSPNTPPKVNNKRELANTPRRVVEEDIEINP- 928
QY 867 SRIAE 871
DB 929 -RMD 932

Search completed: April 26, 2002, 16:19:22
Job time: 352 sec

Mon Apr 29 08:35:39 2002

us-09-556-246-1_copy_200_1263.rpr

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:19:35 ; Search time 47.4 seconds

(without alignments)
823.025 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1263

Sequence: 1 VKDNKKRRKKKPTPKPPVY.....VVALSTAKYKMPESYVFEK 1064

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1168	20.5	5179	1 MUC2_HUMAN	002817 homo sapien
2	950	16.7	1664	1 SLPI_CLODM	006852 clostridium
3	789.5	13.9	1367	1 AMYH_YEAST	P08640 saccharomyc
4	651	11.4	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
5	556.5	9.8	875	1 FPI_MYTEP	Q25460 mytilus edu
6	535.5	9.7	2700	1 ZAN_HUMAN	Q9Y453 homo sapien
7	531	9.7	620	1 EXTN_TOBAC	P13983 nicotiana t
8	533	9.4	1087	1 NEH_MOUSE	P13983 mus musculu
9	530.5	9.3	1162	1 TCNA_TRYCR	P23253 trypanosoma
10	530	9.3	865	1 CPN_DROME	Q02910 drosophila
11	518.5	9.1	872	1 FPI_MYTGO	Q25434 mytilus cor
12	509.5	8.9	662	1 MUC1_XENLA	Q05049 xenopus lae
13	503.5	8.8	1970	1 RPBL_HUMAN	P24928 homo sapien
14	498.5	8.7	1970	1 RPBL_MOUSE	P08875 mus musculu
15	497	8.7	831	1 NEH_RAT	P16884 rattus norv
16	493.5	8.6	467	1 RPBL_CRIGR	P11414 cricetus
17	488.5	8.6	836	1 SSP2_PLATO	Q01443 plasmodium
18	475.5	8.6	1020	1 NEH_HUMAN	P12036 homo sapien
19	475.5	8.3	267	1 EXTN_MAIZE	P14918 zea mays (m
20	471.5	8.3	5376	1 ZAN_MOUSE	P08879 mus musculu
21	468.5	8.2	634	1 HMP1_CANAL	P46593 candida alb
22	467	8.2	817	1 VRP1_YEAST	P37370 saccharomyc
23	463	8.1	2142	1 BAT2_HUMAN	P48614 homo sapien
24	454	8.0	1797	1 VGLX_HSVB	P28968 equine herp
25	448.5	8.0	1161	1 YJ9P_YEAST	P47179 saccharomyc
26	448.5	7.9	670	1 VGS0_HSV1	Q00130 ictalid h
27	442.5	7.8	751	1 FPI_MYTGA	Q27409 mytilus gal
28	439.5	7.7	1083	1 T2D3_HUMAN	Q02468 homo sapien
29	432	7.6	439	1 XP2_XENLA	P17457 xenopus lae
30	432	7.6	3164	1 TEGU_HSV1	P10220 herpes simp
31	426	7.5	2715	1 TRX2_HUMAN	Q9um66 homo sapien
32	424.5	7.4	1125	1 MAP4_MOUSE	P27546 mus musculu
33	424	7.4	307	1 SGS3_DROME	P02840 drosophila

ALIGNMENTS

RESULT	ID	Sequence	Standard	PRT	5179 AA.
1	MUC2_HUMAN	002817, Q14878, 01-JUN-1994 (Rel. 29, Created)			
AC	002817, Q14878, 01-JUN-1994 (Rel. 29, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).				
GN	MUC2 OR SMUC.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Intestine;				
RX	MEDLINE=94132002; PubMed=8300571;				
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;				
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.				
RT	Identification of the amino terminus and overall sequence similarity				
RT	to prepro-von Willebrand factor ";				
RL	J. Biol. Chem. 269:2440-2446(1994).				
RN	[2]				
RP	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.				
RC	TISSUE=Colon;				
RX	MEDLINE=93016075; PubMed=1400449;				
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,				
RT	Kim Y.S.;				
RT	"The human MUC2 intestinal mucin has cysteine-rich subdomains located				
RT	both upstream and downstream of its central repetitive region.";				
RL	J. Biol. Chem. 267:21375-21383(1992).				
RN	[3]				
RP	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.				
RX	MEDLINE=91358717; PubMed=1865763;				
RA	Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,				
RT	Petersen G.W., Kim Y.S.;				
RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays				
RT	and polymorphism.";				
RL	J. Clin. Invest. 88:1005-1013(1991).				
CC	- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND				
CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A				
CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS				
CC	AGENTS AT MUCOSAL SURFACES.				
CC	- SUBUNIT: MULTIMERIC.				
CC	- SUBCELLULAR LOCATION: SECRETED.				
CC	- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,				
CC	BRONCHUS, CERVIX AND GALL BLADDER.				
CC	- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR				
CC	INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).				
CC	- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND				
CC	- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT				
CC	OF SILKWORM HEMOCYTIN.				
CC	- SIMILARITY: CONTAINS 2 WMFC DOMAINS.				
CC	- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).				

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; L21998; AAB95295.1; -
 CC EMBL; M74027; AAB59875.1; -
 CC EMBL; M94131; AAB59163.1; -
 CC EMBL; M94132; AAB59164.1; -
 CC MIM; 158370; -
 CC InterPro; IPR000359; Cys_knot.
 CC InterPro; IPR000561; EGF-like.
 CC InterPro; IPR002400; GF_cys_knot.
 CC InterPro; IPR001007; VMEC.
 CC InterPro; IPR001846; Vwd.
 CC Pfam; PF00007; Cys_knot; 1.
 CC PRINTS; PRO0094; GF_CYS_KNOT.
 CC SMART; SM00214; VMC; 2.
 CC SMART; SM00011; VMC_def; 2.
 CC SMART; SM00216; VWD; 4.
 CC PROSITE; PS00022; EGF_1; UNKNOW_N_1.
 CC PROSITE; PS01185; CTCK_1; 1.
 CC PROSITE; PS01225; CTCK_2; 1.
 CC PROSITE; PS01208; VMEC; 2.
 CC Glycoprotein; Repeat; signal.
 CC SIGNAL; 1 20
 CC CHAIN; 21 5179
 CC DOMAIN; 1401 1747
 CC REPEAT; 1401 1416
 CC REPEAT; 1417 1432
 CC REPEAT; 1433 1448
 CC REPEAT; 1449 1464
 CC REPEAT; 1465 1471
 CC REPEAT; 1472 1478
 CC REPEAT; 1479 1494
 CC REPEAT; 1495 1517
 CC REPEAT; 1518 1533
 CC REPEAT; 1534 1556
 CC REPEAT; 1557 1572
 CC REPEAT; 1573 1596
 CC REPEAT; 1597 1612
 CC REPEAT; 1613 1631
 CC REPEAT; 1636 1651
 CC REPEAT; 1652 1675
 CC REPEAT; 1676 1683
 CC REPEAT; 1684 1699
 CC REPEAT; 1700 1715
 CC REPEAT; 1716 1731
 CC REPEAT; 1732 1747
 CC DOMAIN; 4815 4886
 CC DOMAIN; 4924 4991
 CC DOMAIN; 5075 5160
 CC DISULFID; 5088 5136
 CC DISULFID; 5088 5136
 CC DISULFID; 5088 5136
 CC DISULFID; 5102 5154
 CC DISULFID; 5102 5159
 CC CARBOHYD; 163 163
 CC CARBOHYD; 423 423
 CC CARBOHYD; 670 670
 CC CARBOHYD; 770 770
 CC CARBOHYD; 894 894
 CC CARBOHYD; 1139 1139
 CC CARBOHYD; 1154 1154
 CC CARBOHYD; 1215 1215
 CC CARBOHYD; 1230 1230
 CC CARBOHYD; 1246 1246
 CC CARBOHYD; 1787 1787
 CC CARBOHYD; 1820 1820

FT CARBOHYD 4339 4339 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4351 4351 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4362 4362 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4373 4373 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4422 4422 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4438 4438 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4502 4502 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4616 4616 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4627 4627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4752 4752 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4787 4787 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4881 4881 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4888 4888 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4955 4955 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4970 4970 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5019 5019 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5038 5038 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5069 5069 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1351 1351 H -> L (IN REF. 3).
 FT CONFLICT 1412 1412 L -> S (IN REF. 3).
 FT CONFLICT 1449 1449 M -> T (IN REF. 3).
 FT CONFLICT 1504 1504 G -> S (IN REF. 2).
 FT CONFLICT 4192 4192
 SQ SEQUENCE 5179 AA: 540295 MW: 85CD7571FB9A5663 CRC64:

Query Match 20.5%: Score 1168; DB 1: Length 5179;
 Best Local Similarity 32.5%; Pred. No. 1.9e-40;
 Matches 363; Conservative 53; Mismatches 448; Indels 254; Gaps 40;

QY 21 DEAGSLDNGDFK-----VTPDSTTQH-NKYSTSK-----52
 DB 1312 DHPSSGSDGDREPDVCGAEDIECSRVKDPHLRLQHGOKVQCDVSVGFCKNEDQF 1371
 QY 53 -----TTAKPINRPSLPNSDTSKETSJYNKETTETK 88
 DB 1372 GNGPGLCYDYKIRVNCMPMDKCTTPSSPTTTTLPLPPTTSSPTT-----1427
 QY 89 ETTTINKOTSDGKEKTSKAKETOSIEKTSADLAPTSKVLAKPYKAEYTTKGALATP 148
 DB 1428 -TTTTTPPTTSPPTTTTTP-----LPTT---TSPPISTTTTTPPTTTP 1470
 QY 149 KEPT-----PTPKPEASTPKKEPTTTIKSAP--TTPKEP-----APTTKSAP--TTP 194
 DB 1471 SPTTTPSPPTTTPSPPTTTTTPPTTTPSPPTTTPPTTTPASTTTLPLPPTTSSPTTTP 1530
 QY 195 KEPAPTTKEPAPTTTPKEP-----APTKEPAPTTKSAP--TTPKEPAPTTKKPA 245
 DB 1531 TTPPTTTPSPPTTTPITPPTSTTLPLPPTTSSPTTTPPTTTPPTTTPPTTTPSPPT 1590
 QY 246 PTPKPEAPPTPKPEPTTPPKPEAPPTPKPEAPPTTKEP-----APTAKKPAPTTKE 298
 DB 1591 ITTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1650
 QY 299 PAPTTPKEAPPTTTPKEPSTTTPKEAPPTTKSAPTTPKPAPTTKSAP--TTPKEPSP 355
 DB 1651 PPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1709
 QY 356 TTPKPEAPPTPKPEAPPTPKKPAPTTTPKEAPPTTTPKPAPTTTPKPAPTTTPKPA 414
 DB 1710 TTTTSSSTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1769
 QY 415 ----KEPATTP-----KKLTP 427
 DB 1770 FSPFSTTTPTPPCVPLCNMTGWLDSGRNPHKRGDTLIGDVGCGPMANISCRATMP 1829
 QY 428 -----TTPKEAPTT 437
 DB 1830 DVPDGLGQTVQCDVSVGLCKNEDQKPGCVIPMACLVETINVOGCECVQTPTMTTTP 1889
 QY 438 PEKPAPTTEELAPTTPEEPDTP--TPEEPAPTTPKAADNTPEKAPPTTKEP-----AP 491
 DB 1890 TENPTPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1949

DB 967 VPTTTTSTSTTTTTCSTGNSAGETSGCSPKTTTTCSTSPSETA-----1021
OY 812 ATNSKATTPPKPKP-----TKAPKKPTTKPKMPRRKKTKPTPR 854
DB 1022 ---SESTTSPPTPTVTVSTVTVTEYSTPFGGTTTFTYTKNPTTYLTTP-1077
OY 855 KMTSTPELNPISRIAEAMKQTTTRPNQTPNSKLVENPKSEDAGAGEGTPMLLRPHY 914
DB 1078 -SVTVTNFPTT-ITTVCTSTG-----NSAGETSGCSPKTVTTTPCST 1122
OY 915 FMEVPPDDYLPVPNOGIINPMLSDEFNICNGKPVGTLTLRNGTLVAFRGHYFMKL 974
DB 1123 GICEYTTTATLTVTAVTTVTSTSTGNSA-GKTTGTYTKRSVPT-----YVTVL 1175
OY 975 SPSPSPARRITEWVGIPSPIDTVETRC-----NCEGKT 1009
DB 1176 APSAPVTPATN-----AVPTTITTT-TECSAATNAGET 1207

RESULT 4
MUC1_HUMAN STANDARD; PRT: 1255 AA.

AC P15941; P15942; P13931; P17626; Q14128; Q16442; Q16437; Q9Y4J2;
DT 01-JAN-1990 (rel. 13, Created)
DT 01-APR-1990 (rel. 14, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)
DE MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM) (PEMT)
DE (EPITHELIAL) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)
DE (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H2TAG) (PEANIT-
DE REACTIVE URINARY MUCIN) (PUM) (BREAST CARCINOMA-ASSOCIATED ANTIGEN
DE DF3).
GN MUC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE-pancreas;
RC TISSUE-pancreas;
RX MEDLINE=90368716; PubMed=2394722;
RA Lan M.S., Bactra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL J. Biol. Chem. 265:15294-15299(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90202794; PubMed=2318825;
RA Lichtenberg M.J.T., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT "Epistatin, a carcinoma-associated mucin, is generated by a
RT polymorphic gene encoding splice variants with alternative amino
RT termini".
RL J. Biol. Chem. 265:5573-5578(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-Breast carcinoma;
RX MEDLINE=90368715; PubMed=1697589;
RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RT Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of human tumor-associated
RT polymorphic epithelial mucin.";
RL J. Biol. Chem. 265:15286-15293(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91097524; PubMed=2268309;
RA Lancaster C.A., Peat N., Duhig T., Wilson D.,
RT Taylor-Papadimitriou J., Gendler S.J.;
RT "Structure and expression of the human polymorphic epithelial mucin
RL gene: an expressed VNTR unit.";
RN Blochem. Biophys. Res. Commun. 173:1019-1029(1990).
RP [5]
RC TISSUE-Breast carcinoma;
RX MEDLINE=90276413; PubMed=2351132;

RA Wreschener D.H., Hareuveni M., Tsarfaty I., Smorodinsky N.,
RA Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,
RA Keydar I.;
RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms.";
RL Eur. J. Biochem. 189:463-473(1990).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE-Breast carcinoma;
RX MEDLINE=90276414; PubMed=2112460;
RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horov J.,
RA Zilhan S., Weiss M., Green S., Lathe R., Keydar I., Wreschener D.H.;
RT "A transcribed gene, containing a variable number of tandem repeats,
RT codes for a human epithelial tumor antigen. cDNA cloning, expression
RT of the transfected gene and over-expression in breast cancer
RT tissue.";
RL Eur. J. Biochem. 189:475-486(1990).
RN [7]
RP SEQUENCE FROM N.A.
RX MEDLINE=91033045; PubMed=1688329;
RA Tsarfaty I., Hareuveni M., Horov J., Zaretsky J., Weiss M.,
RA Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschener D.H.;
RT "Isolation and characterization of an expressed hypervariable gene
RT coding for a breast-cancer-associated antigen.";
RL Gene 93:313-318(1990).
RN [8]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=88330762; PubMed=3417635;
RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA Burchell J.;
RT "A highly immunogenic region of a human polymorphic epithelial mucin
RT expressed by carcinomas is made up of tandem repeats.";
RL J. Biol. Chem. 263:12820-12823(1988).
RN [9]
RP SEQUENCE OF 1-169 FROM N.A.
RX MEDLINE=90088473; PubMed=2597151;
RA Abe M., Siddiqui J., Kuft D.;
RT "Sequence analysis of the 5' region of the human DF3 breast
RT carcinoma-associated antigen gene.";
RL Blochem. Biophys. Res. Commun. 165:644-649(1989).
RN [10]
RP SEQUENCE OF 1-109 FROM N.A.
RC TISSUE-Thyroid;
RX MEDLINE=96183746; PubMed=8608966;
RA Weiss M., Baruch A., Keydar I., Wreschener D.H.;
RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT transcriptase polymerase chain reaction of the MUC1 gene.";
RL Int. J. Cancer 66:55-59(1996).
RN [11]
RP SEQUENCE OF 1-89 FROM N.A.
RC TISSUE-Lung;
RX MEDLINE=96181716; PubMed=8604237;
RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
RA Lee L.N., Lun K.T., Wu C.W.;
RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
RT tissues.";
RL Oncology 53:118-126(1996).
RN [12]
RP SEQUENCE OF 1-46 FROM N.A.
RC TISSUE-Breast carcinoma;
RX MEDLINE=90276415; PubMed=2112460;
RA Buluwela L., Liu Q., Lugmani Y.A., Gomm J.J., Coombes R.C.;
RT Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC CYTOSKELETON.
CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM
CC IS ALSO PRODUCED.
CC - ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE
CC SPLICING.
CC - TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL
CC TUMORS, SUCH AS BREAST CANCER.
CC - PTM: HIGHLY GLYCOSYLATED (N- AND O-LINKED CARBOHYDRATES AND SULFIC
CC ACID).
CC - POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT

VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.

-1- SIMILARITY: CONTAINS 1 SEA DOMAIN.

This SWISS-PORT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).

CC EMBL; J05582; AAA60019.1; -

DR EMBL; M32738; AAA35804.1; -

DR EMBL; M32739; AAA35806.1; -

DR EMBL; J05581; AAA59876.1; -

DR EMBL; M61170; AAB53150.1; -

DR EMBL; X52229; CAAB36478.1; ALT_SEQ.

DR EMBL; X52228; CAAB36477.1; ALT_SEQ.

DR EMBL; M35093; AAB59612.1; ALT_SEQ.

DR EMBL; Z17324; CAAB8972.1; -

DR EMBL; Z17325; CAAB8973.1; -

DR EMBL; M31823; AAA35757.1; -

DR EMBL; S81781; AAD14376.1; ALT_INT.

DR EMBL; S81736; AAD14369.1; ALT_INT.

DR EMBL; M21668; AAA59874.1; ALT_SEQ.

DR PIR; A35175; A35175.

DR PIR; B35175; B35175.

DR PIR; S10218; S10218.

DR GLycoSuiteDB; P15941; -

DR MIM; M158340; -

DR MIM; M13720; -

DR InterPro; IPR000082; SEA.

DR Pfam; PF01390; SEA; 1.

DR SMART; SM00200; SEA; 1.

DR PROSITE; PS50024; SEA; 1.

KW Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;

KW Repeat; Alternative splicing.

FT SIGNAL 1 23

FT CHAIN 24 1255

FT DOMAIN 24 1162

FT TRANSMEM 1163 1186

FT DOMAIN 1187 1255

FT DOMAIN 81 960

FT DOMAIN 1034 1151

FT CARBOHYD 957 957

FT CARBOHYD 975 975

FT CARBOHYD 1029 1029

FT CARBOHYD 1055 1055

FT CARBOHYD 1133 1133

FT VARSPPLIC 19 19

FT VARSPPLIC 20 22

FT VARSPPLIC 20 31

FT VARSPPLIC 126 905

FT VARSPPLIC 1077 1087

FT VARSPPLIC 1088 1255

FT CONFLICT 2 2

FT CONFLICT 134 134

FT CONFLICT 154 154

FT CONFLICT 1021 1021

FT CONFLICT 1251 1251

SO SEQUENCE 1255 AA; 122072 MW; 5E28DFC4DE7D9A82 CRC64;

Query Match 11.4%; Score 651; DB 1; Length 1255;

Best Local Similarity 27.8%; Pred. No. 4.4e-20;

Matches 295; Conservative 72; Mismatches 460; Indels 234; Gaps 51

```

QY      75 TSLPYNKETVETKETTNTNKQSTDCAEKTTSAKETQSIEKTSKD-LAPTSKVLAKPT 133
      1111      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      16 TVLPV-----VTGSGHASTPGEKETSAIQRSSPSTENAVASMTSSVLSHS 65

```

[illegible]

RESULT	5	
FPI_MYTED		
ID	FPI_MYTED	STANDARD;
		PRT;
		875 AA

AC 025460:
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ADHESIVE PLACOE MATRIX PROTEIN (POLYPHENOLIC ADHESIVE PROTEIN) (FOOT
 DE PROTEIN 1) (MEFPI) (FRAGMENT).
 GN FPI.
 OS Mytilus edulis (Blue mussel).
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidae;
 OC Mytiloidae; Mytilidae; Mytilus.
 OX NCBI_Taxid=6550;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91025829; PubMed=1367451;
 RA Filipula D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.;
 RT "Structural and functional repetition in a marine mussel adhesive
 RT protein.";
 RL Biotechnol. Prog. 6:171-177(1990).
 RN [2]
 RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.
 RX MEDLINE=83135732; PubMed=6298211;
 RA Waite J.H.;
 RT "Evidence for a repeating 3,4-dihydroxyphenylalanine- and
 RT hydroxyproline-containing decapeptide in the adhesive protein of the
 RT mussel, Mytilus edulis L.";
 RL J. Biol. Chem. 258:2911-2915(1983).
 CC -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
 CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
 CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
 CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
 CC -1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
 CC -1- PTM: THE DECAPEPTIDE A-K-P-S-Y-P-T-Y-K IS POST-TRANSLATIONALLY
 CC MODIFIED AS FOLLOWS: THE SIXTH AND SEVENTH RESIDUES ARE
 CC HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYHENTHIALANINE
 CC (DOPA) DERIVED FROM TYROSINE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X54422; CAA38294.1; .
 DR InterPro: IPR002964; Adhesive_plaq.
 DR InterPro: IPR002965; P-rich_extensn.
 DR PRINTS: PRO1216; ADHESIVER.
 DR PRINTS: PRO1217; PRICHEXTENS.
 DR Repeat: Hydroxylation.
 FT NON_TER 1
 FT DOMAIN 67 870
 FT TANDEN REPEATS OF Y-K-[PA]-K-[LP]-[ST]-Y-
 FT P-P-[ST].
 SQ SEQUENCE 875 AA; 100412 MW; 6EA85312748CAACE CRC64;

Query Match 9.88; Score 556.5; DB 1; Length 875;
 Best Local Similarity 28.7%; Pred. No. 2.1e-16;

Matches 272; Conservative 106; Mismatches 375; Indels 195; Gaps 56;

QY 42 TONKVVSTPKITAKINP-----RP--SLPNSDTSKETSLTVKKEETVET 87
 DB 1 TKHEPVYKPKKTSISAYKPPYQPLKKVDYRPTKSYPTYG-SKNNYLP.LKKKLSYK 59
 QY 88 KETTTNKGSTDSG--KEKTT--SAKETOSIKTSKADLAPTSKVLAKTTPPAETTKG 142
 DB 60 PIKTYVAKNTNPPVYKPKMTYPTTKPKPSYPTTKSKRTYPKKITIYPTTKAKSY-- 117
 QY 143 PAITTKREPPTPKPKE-----PASTTKREPPTTKISAPTT-----KEPAPTTKSAPTT 193
 DB 118 PSSYKPKKTYPTTKPLVPTTKPKPSYPTTKPKPSYPTTKKTYPTSSYKAKPSY 177

QY 194 KEPAPTTKKEPA--PTTKKEPA-----PTTKKEPA--PTTKSAPTTK--EPAPTT 240
 DB 178 P-----PTYKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPY 233
 QY 241 PKKPA--PTTKKEPA--PTTKKEPA--PTTKKEPA--PTTKKEPA--PTTKKEPA--PTAK 289
 DB 234 KAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPY 293
 QY 290 KPA--PTTKKEPA--PTTKKEPA--PTTKKEPS--PTTKKEPA--PTTKSAPTTKKEPA-- 339
 DB 294 KPSYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTK 353
 QY 340 PTTKSAPTTKKEPSPTTKKEPAPTTKKEPAPTTKKEPA--PTTKKEPA--PTTKKEPA-- 395
 DB 354 PSTYKAKSPY--PTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPY 409
 QY 396 TTTKPAPTAKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPA 455
 DB 410 PPTTKAKPTTKAKP-----TYPST-YKAKSPY--PSYKAKSPYPTTKAKSPYPTTKAK 457
 QY 456 EPTTKKEPAPTTKKAAPTTPKEPAPTTKKEPAPTTKKEPA--PTTKKEPA--PTTKKEPA-- 509
 DB 458 KPT-----YPTTKA--KPSYKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPY 506
 QY 510 PKGTAPTTKKEPAPTTKKAPKELAPTTTKKEPTSTSDKAPPTTKGTAPTT--PKE 565
 DB 507 PKLYKPTTK-PKPSYPTTKPKTYPTTK-----PKISYPTTKAKPSY 551
 QY 566 PAPTTPKEPAPTTKPGT--APTTKKEPA--PTTKKPA--PKELAPTTTKG--PTSTSD 617
 DB 552 PATYKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPY 611
 QY 618 KPAPTTKKEPAPTTKKEPA--PTTKKPA--PTT--PETTPPTTSVSPPTTKKE-- 666
 DB 612 KAKSPY--PTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPY 667
 QY 667 PTTHKSPDESTPELSAEPKALENSPK-----EPGPTT--KTPA-----ATKPEMTT 715
 DB 668 PPT-YKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPYPTTKAKSPY 726
 QY 716 AKDPTTTRDITPETTTAAPTMTKETATTTTEKTESKITATTTQVSTTQDTTPKIT 775
 DB 727 YKAPTTKAKPTVPTTKAP--TYKAKPTTKYPTTKAK-----PSYPTTKPKPSYPT 777
 QY 776 TLTITTLAPVTTTKITITTEIMNKPEETAKPKDRATNSKATTPKOKPTKAPK--P 832
 DB 778 TYKSKSYPTTKPKTYPTTK--YKPLTYPTTK-----PKPSYPTTKPKTYPT 826
 QY 833 TSTKPKPTMPVRKPKPTTP-----RKMTSMPDELNPTSR 868
 DB 827 STYKLRKSPYPTTKSKTSYPTTKKISTYSSYKAKTSYPAVKPTNR 874

RESULT 6
 ZAN_HUMAN
 ID ZAN_HUMAN STANDARD: PRT; 2700 AA.
 AC 09Y493; 000218;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ZONADHESIN (FRAGMENT).
 GN ZAN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE OF 1-2379 FROM N.A.
 RX MEDLINE=99018118; PubMed=9799793;
 RA Glocker G., Scherer S., Schatkevooy R., Boright A., Weber J.,
 RA Tsui L.C., Rosenthal A.;
 RT "Large-scale sequencing of two regions in human chromosome 7q22:

RT analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci
 RT reveals 17 genes.
 RL Genome Res. 8:1060-1073(1998).
 RN [2] SEQUENCE OF 2338-2700 FROM N.A.
 RP TISSUE-Testis; PubMed=9126492;
 RC MEDLINE=97271566; Garbers D.L.;
 RA Gao Z., Harumi T., Garbers D.L.;
 RT "Chromosome localization of the mouse zonadhesin gene and the human
 zonadhesin gene (ZAN)."
 RL Genomics 41:119-122(1997).
 CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
 CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
 CC SIGNALING.
 CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
 CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
 CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
 CC ZONA PELLUCIDA.
 CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
 CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
 CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVULOCYCLIC ISTHUS.
 CC -1- DOMAIN: THE WMP DOMAINS 2 AND 3 MAY MEDIATE COVALENT
 CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
 CC -1- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.
 CC -1- SIMILARITY: CONTAINS 4.5 WMP DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.slb.ch/announce/>
 CC or send an email to license@isb.slb.ch).
 CC -----
 CC EMBL: AF053356; AAC78790.1; -
 CC EMBL: U83191; AAC51208.1; -
 DR MIM: 602372; -
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000998; MAM.
 DR InterPro: IPR002965; P_Fich_extensn.
 DR InterPro: IPR002919; TIL.
 DR InterPro: IPR003328; TILA.
 DR InterPro: IPR001007; VMGC.
 DR InterPro: IPR001846; Vwd.
 DR Pfam: PF00629; MAM; 3.
 DR Pfam: PF01826; TIL; 4.
 DR Pfam: PF02345; TILA; 4.
 DR Pfam: PF00094; vwd; 4.
 DR PRINTS: PR01217; PRICHEXTENS.
 DR SMART: SM00137; MAM; 2.
 DR SMART: SM00214; VWC; 1.
 DR SMART: SM00011; VWC_def; 3.
 DR SMART: SM00216; VWD; 3.
 DR PROSITE: PS01186; EGF_2; 3.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS50060; MAM_2; 4.
 DR PROSITE: PS50060; MAM_2; 4.
 KW Glycoprotein; Transmembrane; Cell adhesion; Repeat.
 FT NON_TER 1
 FT DOMAIN <1 109 MAM 1.
 FT DOMAIN 112 136 MAM 2 (PARTIAL).
 FT DOMAIN 161 326 MAM 3.
 FT DOMAIN 322 446 MAM 4.
 FT DOMAIN 483 951 66 x HEPTAPEPTIDE REPEATS (APPROXIMATE)
 FT DOMAIN 953 1065 (MUCIN-LIKE DOMAIN).
 FT DOMAIN 1066 1454 WMPD 1 (PARTIAL).
 FT DOMAIN 1455 1861 WMPD 2.
 FT DOMAIN 1862 2292 WMPD 3.
 FT DOMAIN 2293 2684 WMPD 4.
 FT DOMAIN ? 2684 WMPD 5.
 FT DOMAIN ? EGF-LIKE.
 FT CARBOHYD 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 74

FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1023 1023 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1099 1099 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1618 1618 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1737 1737 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1832 1832 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1878 1878 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2136 2136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2505 2505 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2374 2379 NOKMA -> RACPGP (IN REF. 1).
 FT NON_TER 2700
 FT SEQUENCE 2700 AA: 293013 MW: 80E60C0B12277B1 CRC64;
 SO
 Query Match 9.7%; Score 555.5; DB 1; Length 2700;
 Best Local Similarity 32.6%; Pred. No 5,9e-16;
 Matches 199; Conservative 64; Mismatches 240; Indels 107; Gaps 32;
 QY 58 PINRPSLIPNSDTSKETSILVNKRETYETKETTNNKQTSIDGKERTTSKETSIEKT 117
 DB 445 PVKVLPELPVSPVSS-----TGPSETTGLFENPTISTK-----KPVSTIEKP 487
 QY 118 SAKDLAPTSKVLAPTPKAEITTTGPAITTPKEPTTPPKPASTTPKEPT-----PTTI 172
 DB 488 SVTTEKPT-----VPKEKPTIPTKEPTISTEK-----PTIPSEKPMPSKPTIPSEKPTIL 539
 QY 173 KSAPTPKEPAPTTTKSAPTPKEPAPTTKEPAPTTKEPAPTTKEPAPTTKSAPT 231
 DB 540 TEKPTIPEE-KPTIPESEKPTISTEKPTIPEE--PTIPEETTTTMEEPVAPTEKPSIPT 596
 QY 232 TPKEPAPTPPKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTKEPAPTTKPAATPKP 291
 DB 597 --EKPSIPEK---PTISMEETIISTEKPTICPEKPTIPEK---PTIPEKTSISPEK- 647
 QY 292 APTTPKEPAPTTPKPAPTTTKEPSTTPPKPAPTTTKSAPTTPKEPAPTTKSAPTTPK 351
 DB 648 -PTTPTE-KPTIPEKPTISTEKPTIPEK-PTISPEKLTIPTEKLTIP--EKPTIPT 701
 QY 352 EPSPTTPKEPAPTTTPKEPAPTTPKPA-----PTTPKEPAPTTKEPAPTTKPAATPAP 406
 DB 702 EKPTISTEE--PTTPTEETTTISTEKPSIPMEKPTIPEETTTTSVETIISTEKLTIPW-- 757
 QY 407 KEPAATTPKEPAPT-----TEKKLTIPTEKLA-----PTTPKEPAPTTPELAATPTEE 456
 DB 758 EKPTISTEKPTIPEKPTISPEKLTIP--EKLTIPTEKPTIPEETTTISTEKL--TIPTE 814
 QY 457 PPTTPPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKEPAPTTKPAATTPKGAAPT 516
 DB 815 KPTISPEKPTISTEK-----PTIPE-KPTIPE-----ETTISTEKLTIPT 855
 QY 517 TLKEPAPTTPKPAPELAATTPTEKPTSTT-----SDKPAATTPKGAATTPPKPAATTT 570
 DB 856 --EKPTISPEKLTIPTEKPTISTEKPTIPEKLTIPTEKLTIPTEKPTIPT--EKLTAIR 911
 QY 571 PKEPAPTTKGAATTPKPAATTPKPAPELAATTTKGAATTPKPAATTPKET--A 628
 DB 912 PPHPSPTA-TGLAALVMSHPASTPTV--ILGTTTSSRSSTGMSCP-PAARESCAC 966
 QY 629 PTTTPKEPAPT 638
 DB 967 PASCKSPRS 976
 RESULT 7
 EXTN TOBAC STANDARD; PRT; 620 AA.
 AC P13983;
 DT 01-JAN-1990 (rel. 13, Created)
 DT 01-JAN-1990 (rel. 13, Last sequence update)
 DT 01-MAR-1992 (rel. 21, Last annotation update)
 DE EXTENSIN PRECURSOR (CELL WALL, HYDROXYPROLINE-RICH GLYCOPROTEIN).
 GN HRPNT3
 OS Nicotiana tabacum (Common tobacco).

CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 CC NCB1_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV_XANTHI; TISSUE=Leaf;
 RX MEDLINE=90128263; PubMed=2612909;
 RA Keller B., Lamb C.J.;
 RT "Specific expression of a novel cell wall hydroxyproline-rich
 RT glycoprotein gene in lateral root initiation.";
 RL Genes Dev. 3:1639-1646(1989).
 CC -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
 CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
 CC MAIN ROOT.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
 CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
 CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
 CC GLYCOSYLATED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X13865; CAA32090.1; -
 CC PIR: S06733; S06733;
 CC DR Repeat: Cell wall; Glycoprotein; Signal; Structural protein;
 CC KM Hydroxylation.
 CC FT SIGNAL 1 ?
 CC FT CHAIN 1 ?
 CC FT REPEAT 70 73 EXTENSIN.
 CC FT REPEAT 148 151 H-A-P-P-P.
 CC FT DOMAIN 228 242 H-A-P-P-P.
 CC FT REPEAT 228 235 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
 CC FT REPEAT 236 242 1.
 CC FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.
 CC FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
 CC SEQUENCE 620 AA: 65406 MW: 641DD2278AB28524 CRC64:
 SQ

Query Match 9.7%; Score 551; DB 1; Length 620;
 Best Local Similarity 27.2%; Pred. No. 2.7e-16;
 Matches 170; Conservative 64; Mismatches 304; Indels 88; Gaps 24;

QY 136 AETTK-SPALTP--KEPTPT-----PKPASTPKK---PTPTTSAPPTTKPEAP 184
 DB 24 AETTTGGYLPPTVSGPPSSIGLSPSADPTTPPSGHNVPSP---RHAPPRHAYPPP 80
 QY 185 TTTKSAPTTKPEP-----APTTTKEPAPT--TPKEPAPTTPKPEAPTTKSAPTPE 235
 DB 81 SHGHLPSPVGGRPHNGHLRPSRGFNPPSPVISPNSHPSPGAPRPSGPHLDSHGCR 140
 QY 236 PAPTPPKKAPR---TPKRAPTTTKEPPTTKPEAPTTK--EPAPTTKPEAPTPAKK 290
 DB 141 P-----PSPSHGNAPSGHTPRGQHPSHRRSPSRGHNHPPRYAOPPTPIYS 193
 QY 291 PAPTPKPEAPTTKPEAPTTKPEAPTTK---EPAPTTKSAPTTPKPEAPTTKS-- 345
 DB 194 PEPV--QPPR--YSPRPHTVQTPSPSPSGHROPRTTHRHADPTTHKADPTHQPSPL 249
 QY 346 ---APTTKPEPSPPTTKPEAPTTKPEAPTTKPKRAPTTTKEPAPT--TPKEPAPTTPK 399
 DB 230 RHLPSRPOPOPTYSPPPAVQSPQSPPTYSPPPTSPSPSPSPSPSPSPSPSPSP 309
 QY 400 KPAPT---APKRAPTTTKETAPTPKL-PTT-----PEKLAPTTPEKAPTTPEEL 449
 DB 310 PPTPTPTSPSPPAVSPPTYSPPPTYLPLPSSPTYSPPPVVSPPPPSYSPPTPTYL 369
 QY 450 APPTPEEPPTTPEPAPTTPKAAP---NTPKPEAPTTTKEPAPTTPKPEAPTTPKETA 506

DB 370 PPPSSPPSPSPSPPTTQSPSPPAVSPPLPAPPTYSPPPT--YSPPTVQAP 427
 QY 507 PTPPKGTAPTTLKEPAPTTPKKPAKELAPTTTKEPSTSDKRAPTPKGAPTTKEP 566
 DB 428 P-----LPPTYSPPPAVSPSPPTTSPSPPTYSPPPAVQAPPPPTYSPPPAVSP 481
 QY 567 APPTPKPEAPTTKGAAPTTLKEPAPTTPKPAKELAPTTTGTGPTSDKRAPTPKE 626
 DB 482 PPSPTYSPPPAVQAPLPPTFSPPPRHILPPPHRQ---RPPTVQGPSPTFSF 538
 QY 627 TAPTTKEPAPT--TPKRAPTTTPTPTPTTSEVSTPTTKEPTTHKSDESPELISAE 684
 DB 539 PPRQHSPPPPHMQPPTPTPTVGGQPPSPPTFSAPPROHSPPPPHQRPPPTVYGP 598
 QY 685 PTPKALNSPEKPGVPTTKPAATKP 710
 DB 599 PSD-----PTTYSPPSPPP 612

RESULT 8
 NFH_MOUSE STANDARD; PRT: 1087 AA.
 AC P19246; 061959;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
 GN NEFH OR NFH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89121513; PubMed=3220257;
 RA Julien J.-P., Cole F., Beaudet L., Sidky M., Flavell D., Grosfeld F.,
 RA Mushynski W.;
 RT "Sequence and structure of the mouse gene coding for the largest
 RT neurofilament subunit.";
 RL Gene 68:307-314(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89089138; PubMed=3145094;
 RA Sheindman P.S., Garden M.J., Lees J.F., Lazzarini R.A.;
 RT "The structure of the largest murine neurofilament protein (NF-H) as
 RT Brain Res. 464:217-231(1988).
 RN [3]
 RC SEQUENCE FROM N.A.
 RC STRAIN=SWISS WEBSTER; TISSUE=Brain;
 RA Carden M.J.;
 RA Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PRT: THERE ARE A NUMBER OF REPEATS OF THE TRIPETIDE K-S-P, NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
 CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M24496; AAA39813.1; JOINED.
 CC EMBL; M23349; AAA39813.1; JOINED.
 CC EMBL; M24495; AAA39813.1; JOINED.
 CC EMBL; M35131; AAA39809.1; ALT_FRAME.
 CC EMBL; M31012; CA83229.1; .
 CC PIR; J03068; QFM5H.
 CC PIR; A43778; A43778.
 CC MGD; MGI:97309; NfH.
 CC InterPro: IPR001664; IF.
 CC Pfam: PF000226; IF; 1.
 CC PROSITE; PS00226; IF; 1.
 CC Intermediate filament; Coiled coil; Neurone; Phosphorylation;
 KW Repeat.
 FT DOMAIN 1 97 HEAD.
 FT DOMAIN 98 408 ROD.
 FT DOMAIN 409 1087 TAIL.
 FT DOMAIN 436 517 GLU-RICH (ACIDIC).
 FT DOMAIN 519 886 50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.
 FT DOMAIN 887 1087 GLU/LYS-RICH.
 FT DOMAIN 98 129 COIL 1A.
 FT DOMAIN 130 141 LINKER 1.
 FT DOMAIN 142 239 COIL 1B.
 FT DOMAIN 240 261 LINKER 12.
 FT DOMAIN 262 283 COIL 2A.
 FT DOMAIN 284 287 LINKER 2.
 FT DOMAIN 288 408 COIL 2B.
 FT CONFLICT 133 133 K -> QA (IN REF. 2 AND 3).
 FT CONFLICT 199 199 A -> AR (IN REF. 2 AND 3).
 FT CONFLICT 281 281 L -> T (IN REF. 2 AND 3).
 FT CONFLICT 492 492 L -> G (IN REF. 2 AND 3).
 FT CONFLICT 551 551 P -> PREAKSP (IN REF. 3).
 FT CONFLICT 689 712 MISSING (IN REF. 3).
 FT CONFLICT 714 714 G -> A (IN REF. 3).
 FT CONFLICT 814 814 V -> M (IN REF. 2 AND 3).
 FT CONFLICT 843 843 T -> N (IN REF. 2 AND 3).
 SO SQUENCE 1087 AA; 116612 MW; 57BAC76A38ED1CB9 CRC64;

Query Match 9.4%; Score 533; DB 1; Length 1087;
 Best local similarity 27.5%; Pred. No. 2,3e-15;
 Matches 216; Conservative 81; Mismatches 330; Indels 158; Gaps 38;

QY 21 DEAGSGINDGDFKVT-----TPDSTGHNKVSSTPKITTKAPINPRSLPPNSDTSKET 75
 Db 404 ECHRGGSPFSLEGLPKIPISIT-HIKVSEEMIKVYE-----KSEKET 449
 QY 76 SLVANKETVETKETTNTKOTSDGKEKTSAKETOSIEKTSADLAFTSKVLAKPTPK 135
 Db 450 VIVSGQEEIIVTGVTEEDKEAQQGEGEAEKEEELAAVSPAEAAASPEKE 509
 QY 136 AETTTGPAITTKPEPTPTTKPEASITPKPTTIKSATTTKEPAFTTKAPPTPK 195
 Db 510 TKSRYKEAKSPGEAKSPGAEKSPA---EAKSPGEAKS-PGEAKSPGEAKSPAEKSPA 564
 QY 196 EP-APTTTKEPAPTTKPEAPTTTKEP---APTTTKSAPTTPKEPA---PTTPKKPA- 245
 Db 565 EPKSPAEKSPA--EPKSPA--TKSPGEAKSPSIAKS-PAEAKSPAEKSPAEKSPA 619
 QY 246 ---PTTPKEPA---PTTPKEP---PTTPKEPAPTTKPEAPTTKPEPA---PTAKK 290
 Db 620 AKSPAEKSPAEKSPATVKSPEAKSPGEAKSPAEKSPA---EAKSPAEKSPAEKSPA 676
 QY 291 P---APTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 342
 Db 677 PGEAKSPAEKSPAEKSPA--KSPA---EAKSPAEKSPAEKSPAEKSPA--PAAVKSPALAK 729

QY 343 TKSAPTTTKE--PSPTTKEPA---PTTPKPEAPTTPKKPAAPTTTKEPAPTTKEPAPT 397
 Db 730 SPAAVKSPGEAKSPGEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 789
 QY 398 TKKPAAPTTKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 457
 Db 790 AKSPYKEDIKKPAEAKSPAEKSPA-KSPVKEGAKPPEKAKPLDVKSPEAQTPVGEATVPDI 848
 QY 458 TPTTDEE-PAPTTKAAAPNTKEPAPT---KPEAPTTTKEPAPTTTKEPAPTTTKEPAPT 505
 Db 849 RP--EEOVKSPEAKSPAEKSPA--KEEAKTSEKAPKKEVKSVEKAEVAKKEPKVDEEK 904
 QY 506 APPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 563
 Db 905 TLPTTKEKESKDEAPPEAKPKVEEKEPTTEKPKNSTAEAKKEEAGEKKAVASE 964
 QY 564 KEPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 623
 Db 965 EEPKAGVKEEA--KPEKETTTTKEADTKAKES---KPETEKP----- 1006
 QY 624 PKETAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 683
 Db 1007 -----KKEMPAPER-----KTKKEKTESKKEP----- 1034
 QY 684 EPTPKALENSPKPEGVPTTKPAATKPEMTTAKODTTERDLRTPTTAAKMTKETA 743
 Db 1035 ---PK-MEAKVKEDDKSLKEP--SKPTEKAEKSSSTQKESQPE-----KTEDEKA 1082
 QY 744 TTKET 748
 Db 1083 TKGEK 1087

RESULT 9
 TCNA_TRYCR STANDARD; PRT; 1162 AA.
 ID TCNA_TRYCR
 AC P23253;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE STALIDASE (EC 3.2.1.18) (NEURAMINIDASE) (NA) (MAJOR SURFACE ANTIGEN).
 GN TCNA.
 OS Trypanosoma cruzi.
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCBI_TaxID=5693;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SLIVTO X-10/4;
 RX MEDLINE=91277609; PubMed=1711561;
 RA Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,
 RA Prioli R.P.;
 RT "The Trypanosoma cruzi neuraminidase contains sequences similar to
 RT bacterial neuraminidases, wtrd repeats of the low density lipoprotein
 RT receptor, and type III modules of fibronectin.";
 RL J. Exp. Med. 174:179-191(1991).
 RN [2]
 SO SUBCELLULAR LOCATION.
 RP MEDLINE=91376547; PubMed=1896773;
 RA Prioli R.P., Mejia J.S., Aji T., Alkawa M., Pereira M.E.A.;
 RT "Trypanosoma cruzi: localization of neuraminidase on the surface of
 RT trypanostigotes.";
 RL Trop. Med. Parasitol. 42:146-150(1991).
 CC -1- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN
 CC PARASITE INVASION OF CELLS.
 CC CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC
 CC LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACETYLNEURAMINYL
 CC RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACETLATED
 CC NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,
 CC GLYCOLIPIDS OR COLONOMIC ACID.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POSSIBLE).
 CC -1- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPANOSTIGOTES, MINIMUM
 CC IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.

CC -1- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT
CC DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TNCA
CC PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.
CC -1- SIMILARITY: CONTAINS 3 BNR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M61732; AAA30255.1; -
DR PIR: JH0557; JH0557.
DR HSSP: P29766; JDLIL.
DR InterPro: IPR002860; BNR.
DR Pfam: PF02012; BNR. 2.
KW Hydrolase: Glycosidase: Glycoprotein; Repeat; GPI-anchor;
KW Phosphorylation.
FT DOMAIN 1 457 CYS-RICH.
FT REPEAT 23 34 BNR 1.
FT REPEAT 163 174 BNR 2.
FT REPEAT 209 220 BNR 3.
FT DOMAIN 458 588 FIBRONECTIN TYPE-III.
FT DOMAIN 589 1120 44 X 12 AA TANDEM REPEATS. LTR DOMAIN.
FT CARBOHYD 342 342 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1162 AA; 120032 MW; 07049221897C6A40 CRC64;

Query Match 9.3%; Score 530.5; DB 1; Length 1162;
Best Local Similarity 29.4%; Pred. No. 3e-15;
Matches 177; Conservative 100; Mismatches 212; Indels 113; Gaps 34;

QY 152 TPTTPKE-PASTTPKEPTTTTSAPTTKEPAATTTKSAPTTKEPAATTTKEPAATTT
DB 599 TPSTPADSSAHSTPTSPVDS--SAHSTPTPADSSAHGTPSPVDSAHGT---PSTP 651
QY 211 KE-PAATTTKEPAATTTKSAPTTKE-PAATTPKP-----APTTPKEPA---PTTP 257
DB 652 ADSAAGTPTSPVDSAHSTPTSPVDSAHSTPTSPVDSAHGAPSTPADSSAHGTPSTP 711
QY 258 KEPTP-TTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPS 316
DB 712 VDSAHGTPSTPADSSAHSTPTPADSS-----AHSTPTPADSSAHSTPTPADSSA 764
QY 317 PTTPKEPAATTTKSAPTTKEPAATTTKSAPTTKEPSPTTKEPAATTTKEPA---PT 372
DB 765 HGTPSTPADSSAHSTPT--PADSSAHGTPSTPVDSAHST---PSTPVDSAHGTPS 817
QY 373 TP-KKPAATTTKEPAATTTKKAAPTARKEPAATTTKEPAATTTKEPAATTTKL---TP 427
DB 818 TPVDSAHSTPTSPVDSAHGTPSTPVDSAHSTPTPADSSAHST--PSTPADSSAHGTP 876
QY 428 TTP-EKLAATTTKEKAPTTPEELAPTTPE---PTTPPEEPA---PTTP-KAAAPT 477
DB 877 STPVDSAHSTPTPADSSAHST--TSTPVDSAHSTPTPADSSAHGTPSTPVDSAHGT 935
QY 478 KPEPAATTTKEPAATTTKEPAATTTKEPAATTTKGAATTTKEPAATTTKKAPELAPT 537
DB 936 PSTPADSS-----AHSTPTPADSSAHST--PSTPADSS-----AHSTPTSPVDSAHST 983
QY 538 TTKEPTSTSDKPAATTTKGAATTTKE-PAATTTKEPAATTTKGAATTTKEPAATTTK 596
DB 984 -----PSTPADSSAHST-----PSTPADSSAHSTPTPADSSAHSTPTSPVDSAHSTPS 1033
QY 597 KPAPELAPTTTKPTSTSDKPAATTTK---ETAPTTTKEPAATTTKKAPELAPT 652
DB 1034 TP-----ADSSAHGTPSTPADSSAHSTPTSPVDSAHSTPTPADSS-----AHGTPSTPA 1084

QY 653 PTSEVSTPTTKEPTTTKHSPTDESP-ELASPTPEKALENSPEKGVPTTTPATTKPE 711
DB 1085 DSSAH-STPTPAD-SSAHGTP--STPADSSAHSTP-----STPADSSAN 1125
QY 712 MT 713
DB 1126 GT 1127
RESULT 10
CPN DROME
ID CPN DROME STANDARD: PRT: 865 AA.
AC 002910;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE CALPHOTIN.
GN CPN OR CAP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CANTON-S;
RA MEDLINE=93165729; PubMed=8094559;
RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.",
RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CANTON-S;
RA MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
RT calcium and contains a leucine zipper".
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
CC REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
CC OF CA+2 PER MOL OF PROTEIN.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPDENSE COMPARTMENT.
CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
CC COMPOUND EYES AND OCELLI.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
CC DEVELOPMENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L02111; AAA28405.1; -
DR EMBL: L05080; AAA28420.1; -
DR PIR: A47282; A47282.
DR Flybase: FBgn0010218; Cpn.
KW Calcium-binding.
FT CALCULUM 36
FT CONFLICT 43 43 A -> T (IN REF. 2).
FT CONFLICT 64 64 I -> V (IN REF. 2).
FT CONFLICT 76 76 T -> A (IN REF. 2).
FT CONFLICT 100 100 P -> PP (IN REF. 2).
FT CONFLICT 126 127 VQ -> AP (IN REF. 2).
FT CONFLICT 154 154 I -> V (IN REF. 2).
FT CONFLICT 160 160 S -> T (IN REF. 2).
FT CONFLICT 534 534 A -> E (IN REF. 2).
FT CONFLICT 699 699 I -> T (IN REF. 2).
FT CONFLICT 703 703 V -> L (IN REF. 2).
FT CONFLICT 721 721 D -> E (IN REF. 2).

SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFE CRC64;

9.38; Score 530; DB 1; Length 865;

Matches 233; Conservative 73; Mismatches 433; Indels 152; Gaps 41;

Matches 23:

[illegible]

DT	30-MAY-2000	(Rel. 39, last sequence update)
DT	20-AUG-2001	(Rel. 40, last annotation update)
DE	ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR (FOOT PROTEIN 1) (MCPPI).	
GN	FPI.	
OS	Mytilus coruscus (sea mussel).	
OC	Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;	
CC	Mytiloidea; Mytilidae; Mytilus.	
OX	NCBI_TaxID=42192;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Foot.	
RX	MEDLINE=96394686; PubMed=8798340;	
RA	Inoue K., Takeuchi Y., Takeyama S., Yamaha E., Yamazaki F., Odo S.,	
RA	Harayama S.;	
RT	"Adhesive protein cDNA sequence of the mussel Mytilus coruscus and	
RT	its evolutionary implications.";	
RL	J. Mol. Evol. 43:348-356(1996).	
CC	-I- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS	
CC	PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S	
CC	ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A	
CC	FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.	
CC	-I- SUBCELLULAR LOCATION: SECRETED.	
CC	-I- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.	
CC	-I- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.	
CC	-I- PTM: PROLINES IN THE REPEAT ARE HYDROXYLATED (SINGLE OR DI-) AND	
CC	ALSO TYROSINE (THUS PRODUCING DOPA = 3,4-DIHYDROXYPHENYLALANINE).	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation at	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see http://www.isb-stb.ch/announce/	
CC	or send an email to license@isb-stb.ch).	
CC	-----	
DR	EMBL, D63777; BAA09850.1; -;	
DR	InterPro: IPR002964; Adhesive_plaq.	
DR	InterPro: IPR002965; P-rich_extensin.	
DR	PRINTS: PR01216; ADHESIVEI.	
DR	PRINTS: PR01217; PRICHEXTENS.	
KW	Signal; Repeat; Hydroxylation.	
FT	SIGNAL	1 20
FT	CHAIN	21 872
FT	DOMAIN	21 41
FT	DOMAIN	124 872
FT	DOMAIN	184 192
FT	DOMAIN	213 221
FT	SEQUENCE	872 AA; 10167 MW; 96CCT0D7C75FF3C4 CRC64;

RESULT	11	
FPL_MYTCO		
FPL_MYTCO	STANDARD;	872 AA.
AC	025434;	
DT	30-MAY-2000 (Rel. 39, Created)	

FT DISULFID 573 599 BY SIMILARITY.
 FT DISULFID 583 598 BY SIMILARITY.
 FT DISULFID 593 610 BY SIMILARITY.
 FT DISULFID 621 647 BY SIMILARITY.
 FT DISULFID 631 646 BY SIMILARITY.
 FT DISULFID 641 658 BY SIMILARITY.
 FT DISULFID 641 658 BY SIMILARITY.
 FT VARIANT 276 276 K -> E.
 FT VARIANT 354 354 C -> R.
 FT VARIANT 415 415 T -> A.
 SQ SEQUENCE 662 AA: 67774 MW: F085277F1ED2ED40 CRC64:

Query Match 8.9%; Score 509.5; DB 1; Length 662;
 Best Local Similarity 28.1%; Pred. No. 1.4e-14;
 Matches 223; Conservative 48; Mismatches 276; Indels 247; Gaps 30;

QY 92 TTNKQTSDDCKEKTSAKESQSIKESIASDAPTSVLAKPPTKATTTTGGALTTTPKRP 151
 DB 3 TTTAAVAATGCKDTTAAEGSAAEKTA-----AGEVSAPPT--AAVAATGEDATT----- 51
 QY 152 TPTTPKEPASTTTPKEPTTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 211
 DB 52 -----AATAAETTAAGAPPTTTAPATTAAGKAPTTAAATVAPTTAAGAPTTAT 103
 QY 212 EPAPTTPKEPAPTTPKS-APTTPKEPAPTTPKAPPTTPKEPAPTTPKEPTTPKEPA- 269
 DB 104 GKAPATTAAPVPTTAASKAPTAAATHTAAATAAPTAAASAASAKSERTSSSEEHCH 163
 QY 270 --PTTPKEAPT--TPKEPAPTAPKAPPTTPK-----EPAPTTPKEPAPTTPKE 314
 DB 164 VKSKRMCGSKGTTKKO-----CKNNCCEDPRGHGSHCFHKKRPGHSHHEHTTTTK- 218
 QY 315 PSPPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTP 374
 DB 219 -----APTIIQIATTTT--TPTTT-----TTTKATPTT----- 245
 QY 375 KKPAPTTPKEPAPTTPKEPAPTTPKAPAPKAPKEPAPTTPKEAPTTPKTLTTPKEKLA 434
 DB 246 -----TTTKAPTTP-----TTTKATPTT-----TPTT----- 270
 QY 435 PTPPEKAPTTPPELAPTTPPEPTTPPEBAPTTPKRAAPNPPKEPAPTTPKEPAPTTP 494
 DB 271 -----TTTKATPTTPPTTTT----- 289
 QY 495 KEPAPTTPKEAPTTPKGTATPTLKEPAPTTPKAPKAPKELAPTTTKEPTSTSDKAPPT 554
 DB 290 -----TTTTKATTTTTTTTSGECKMEPSK-----RDCGSGITTESOCR 328
 QY 555 PKG-----TAPPT-----PKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKAPKELA 604
 DB 329 TKGCCFDSISQTKWCFTYLSQVADKVEPSQRVDCGRGIT-----ADOCROKNCDFDS 384
 QY 605 PTTTGTSTSDKAPAPTTPKEAPTTPKEPAPTTPKAPPTTPETPTTSEVSTPTT 664
 DB 385 ISGTKWCFTSTSOVAA--TKTTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 436
 QY 665 KEPTTIHSPDESPELSAEPITKPALENSPEKPGVITKTPATKPEMTTAKDKTTERD 724
 DB 437 TTTT-----TTTA-----TTTTTTTTTTTTT-----KAT----- 463
 QY 725 LRTPEPTTAAPKMTKETATTEKTESKITATTVOVSTTTTODTTPKITTLLTTLAP 784
 DB 464 -TTTTTTTTT-----TTTTKAT-----TTTTTTTTTTTTT-----KAT----- 510
 QY 785 KYTTTKKITTITTELMNKEEFRAKPRDAINSKATTPKOKOTKAPKKTSTSKRPK-----T 840
 DB 511 TTTTATTTTTTSGECKME-----PSKRADCGPGITTESQCKSGCCFDSISIPQTKWCFTS 566
 QY 841 MPVRKPKTTPTR 854
 DB 567 LPQVADCKVAPSSR 580

RESULT 13
 RPB1_HUMAN
 ID RPB1_HUMAN STANDARD; PRT; 1970 AA.
 AC P24928;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPB1).
 GN POLR2A.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 CC NCBI_TaxID=9606;
 RX MEDLINE=92178992; PubMed=1542581.
 RA Wiltzerlth M., Acker J., Vicaire S., Vigneron M., Keding C.,
 RT "Complete sequence of the human RNA polymerase II largest subunit.",
 RL Nucleic Acids Res. 20:910-910(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=95347616; PubMed=7622068;
 RA Mita K., Tsuji H., Moriyama M., Takahashi E., Neno M.,
 RA Ichimura S., Yamuchi M., Hongo E., Hayashi A.,
 RT "The human gene encoding the largest subunit of RNA polymerase II.",
 RL Gene 159:285-286(1995).
 CC - FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC - CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
 CC RNA(N).
 CC - SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
 CC - SUBCELLULAR LOCATION: NUCLEAR.
 CC - PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
 CC THE PHOSPHORYLATION ACTIVATES POL2.
 CC MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
 CC III FOR 5S AND 5.8S GENES.
 CC - SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X63564; CAA5125.1;
 DR EMBL: X74874; CAA52862.1;
 DR EMBL: X74873; CAA52862.1; JOINED.
 DR EMBL: X74872; CAA52862.1; JOINED.
 DR EMBL: X74871; CAA52862.1; JOINED.
 DR EMBL: X74870; CAA52862.1; JOINED.
 DR PIR: S21054; S21054.
 DR MIM: 180660;
 DR InterPro: IPR000684; RNA_polII_repeat.
 DR InterPro: IPR000722; RNA_pol_A.
 DR InterPro: IPR002879; RNA_pol_A2.
 DR Pfam: PF00623; RNA_pol_A2.
 DR Pfam: PF01854; RNA_pol_A2; 1.
 DR PROSITE: PS00115; RNA_pol_II_REPEAT; 43.
 DR TRANSFERASE: DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
 KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
 FT ZN-FING 71 87
 FT DOMAIN 1590 1958 C2H2-type (POTENTIAL).
 FT CONFLICT 1067 1067 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
 FT CONFLICT 1449 1449 W -> L (IN REF. 2).
 FT CONFLICT 1449 1449 D -> Y (IN REF. 2).
 SQ SEQUENCE 1970 AA: 217205 MW: 6876C25692A657E CRC64;

Query Match

8.8%; Score 503.5; DB 1; Length 1970;

QY 501 TPKETAPTPKGTAPTTKEAPATTPK-KPAKELAPTTKEPTTSDKAPATTPKGT 559
 Db 1829 SP-SYSTSPKXT-----PTSPSYSPSPDETPASPK--YSPTSPKYSPTSPK-YS 1875
 QY 560 PTPPKKAPATTPKAPATTPKGTAPTTKEAPATTPKAPKAPATTPKGT-STSDK 618
 Db 1876 PTPSPK-SPTPPKY-SPSPK--TSPSPSPVTPSPK-----YSPTS--PTSPSPK 1922
 QY 619 PAPTTPKAPATTPKPEP--APTTPKAPATTPETPTPTTKEPTTTHKSPDE 676
 Db 1923 YSTSP-TTSPSPKSTSPSTSPSPK-SPSP-----TSPSTSPAL-----SPD 1965
 QY 677 STPE 680
 Db 1966 SDEE 1969
 RESULT 15
 NFH_RAT STANDARD; PRT: 831 AA.
 AC P16884; 063368;
 DT 01-ANG-1990 (Rel. 15, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 20-ANG-2001 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).
 GN NFH OR NFH.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID:10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RX MEDLINE=89065087; PubMed=3143606;
 RA Breen K.C., Robinson P.A., Wilson D., Anderson B.H.;
 RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
 RT Identification of putative phosphorylation sites.";
 RL FBS Lett. 241:213-218(1988).
 RN [2]
 RP SEQUENCE OF 37-831 FROM N.A.
 RX MEDLINE=88309090; PubMed=2457365;
 RA Daultigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,
 RA Daultigny P.;
 RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and
 RT in situ detection.";
 RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
 RN [3]
 RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.
 RX MEDLINE=87080760; PubMed=2878828;
 RA Robinson P.A., Wilson D., Anderson B.H.;
 RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide
 RT (NF-H).";
 RL FBS Lett. 209:203-205(1986).
 RN [4]
 RP SEQUENCE OF 318-831 FROM N.A.
 RX MEDLINE=89184647; PubMed=2928342;
 RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D.,
 RA Smolowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;
 RT "Cloning of a cDNA encoding the rat high molecular weight
 RT neurofilament peptide (NF-H): developmental and tissue expression in
 RT the rat, and mapping of its human homologue to chromosomes 1 and
 RT 22.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC OBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- THERE ARE A NUMBER OF REPEATS OF THE TRIPLETT K-S-P. NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE

CC OF AXONAL CALIBER.
 CC -1- PTH: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
 CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC -----
 CC EMBL; M37227; AAA41693.1; ALT_FRAME.
 CC EMBL; X13804; CAA32038.1; ALT_FRAME.
 CC EMBL; M21964; AAA41695.1; -.
 CC EMBL; J04517; AAA41692.1; -.
 CC PIR; A30796; A30796.
 CC PIR; A25649; A25649.
 CC PIR; B25649; B25649.
 CC PIR; S02003; S02003.
 CC InterPro: IPR001664; IF.
 CC Pfam: PF00038; filament; 1.
 CC PROSITE; PS00226; IF; 1.
 CC Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.
 CC KW NON_TER 1
 CC FT DOMAIN 1
 CC FT 276 641
 CC FT 164 164
 CC FT 185 185
 CC FT 193 193
 CC FT 199 199
 CC FT 346 346
 CC FT 373 373
 CC FT 482 482
 CC FT 485 485
 CC FT 570 571
 CC FT 591 591
 CC FT 727 727
 CC FT 757 759
 CC FT 769 769
 CC FT 775 775
 CC FT 831 AA; 89486 MW; 1B0973C3F13EF768 CRC64;
 CC SEQUENCE
 CC
 CC Query Match 8.7%; Score 497; DB 1; Length 831;
 CC Best Local Similarity 27.6%; Pred. No. 5.2e-14;
 CC Matches 195; Conservative 59; Mismatches 307; Indels 146; Gaps 35;
 CC
 CC 37 PDSTGQHNKVSPTPKITAKPINRPSLPNSDSKETSILVNNETVETKETTNNKQ 96
 Db 197 PSMST--HIKVKSEKIKYVE-----KSEKEIVVEEQEEIYQVEEVEED 242
 QY 97 TSTDGKKTSAKETOSIEKTSKADLAPTSKYIAKTPPAETTTKPAATTTKEP---T 152
 Db 243 KEAGGEEEEAEAGEEAATTS-----PPAEASAP-----EKEKSPVKEAKSPK 293
 QY 153 PTPPKKPA-STTPKEPTPTTKSAPTTPKPA-----PTTKSAPTTPKPA-----PTTK 203
 Db 294 PAEAKSPAEKSPAEKSPAVAKSPAEVKSAPAEKSPAEKSPAEKSPAEVKSAPATV 352
 QY 204 EP-----APTTPKPA-----PTTKKPA-----PTTKSAPTTPKPA-----PTTKKPA-- 245
 Db 353 SPGAKSPAEKSPAEVKSAPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPATV 411
 QY 246 --PTTPKPA-----PTTKPEP-----PTTKKEAPATTTKEP-----PTA 287
 Db 412 KSPYEAKSPAEVKSAPVAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPV 470
 QY 288 PKKAPATTPKPAATTPKKEAPATTTKEPSPPTTKKPA-----PTTKS-----APTTPKEP 338

[illegible]

Search completed: April 26, 2002, 16:19:50
Job time: 375 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:17:05 ; Search time 49.78 Seconds
(without alignments)
480,986 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1263

Perfect score: 5698

Sequence: 1 VKDKKKRRTKKKPPKPPVY.....VAALSTAKYKNWPEYFFK 1064

Scoring table: BLOSUM62

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: 1: /cgn2.6/ptodata/2/1aa/5A.COMB.pep.*

2: /cgn2.6/ptodata/2/1aa/5B.COMB.pep.*

3: /cgn2.6/ptodata/2/1aa/6A.COMB.pep.*

4: /cgn2.6/ptodata/2/1aa/6B.COMB.pep.*

5: /cgn2.6/ptodata/2/1aa/CTUS.COMB.pep.*

6: /cgn2.6/ptodata/2/1aa/backfilest.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	522	9.2	805 4 US-09-103-429A-4	Sequence 4, Appli
2	508	8.9	1837 3 US-08-928-361B-5	Sequence 5, Appli
3	506.5	8.9	744 6 5202236-25	Patent No. 5202236
4	498.5	8.7	786 4 US-09-103-429A-3	Sequence 3, Appli
5	489	8.6	1721 3 US-08-700-651-5	Sequence 5, Appli
6	488.5	8.6	1721 3 US-08-928-361B-6	Sequence 6, Appli
7	488.5	8.6	826 1 US-07-638-431-2	Sequence 2, Appli
8	488.5	8.6	826 5 PCT-US92-00018-2	Patent No. 5202236
9	476.5	8.4	652 6 5202236-13	Sequence 5, Appli
10	452	7.9	960 4 US-09-219-849-5	Sequence 5, Appli
11	424.5	7.4	1867 2 US-08-479-537A-5	Sequence 5, Appli
12	424.5	7.4	1867 4 US-08-083-116-5	Sequence 5, Appli
13	424.5	7.4	2035 2 US-08-479-537A-2	Sequence 2, Appli
14	424.5	7.4	2035 4 US-09-083-116-2	Sequence 2, Appli
15	419.5	7.4	2476 2 US-08-276-967-2	Sequence 2, Appli
16	419	7.4	1185 4 US-09-041-886-23	Sequence 2, Appli
17	417	7.3	829 1 US-08-642-255-132	Sequence 132, App
18	417	7.3	829 1 US-08-397-633A-53	Sequence 53, Appli
19	417	7.3	837 1 US-08-175-155-68	Sequence 68, Appli
20	417	7.3	837 1 US-08-477-509B-103	Sequence 103, App
21	417	7.3	837 1 US-08-642-255-101	Sequence 75, Appli
22	417	7.3	837 3 US-08-707-237A-75	Sequence 103, App
23	417	7.3	837 3 US-08-482-085B-103	Sequence 103, App
24	417	7.3	897 1 US-08-397-633A-50	Sequence 50, Appli
25	413.5	7.3	907 3 US-08-783-774-2	Sequence 19, Appli
26	413.5	7.3	907 5 PCT-US95-04611A-19	Sequence 65, Appli
27	404.5	7.1	408 1 US-07-609-716-65	Sequence 65, Appli

28	404.5	7.1	408 4 US-08-475-411A-65	Sequence 65, Appli
29	404.5	7.1	408 4 US-08-478-029A-65	Sequence 65, Appli
30	398.5	7.0	682 1 US-08-642-255-126	Sequence 126, App
31	398.5	7.0	682 1 US-08-397-633A-36	Sequence 36, Appli
32	381	6.7	1848 4 US-08-296-791-6	Sequence 6, Appli
33	381	6.7	1848 5 PCT-US95-10661A-6	Sequence 6, Appli
34	376	6.6	960 4 US-09-219-849-6	Sequence 6, Appli
35	374.5	6.6	1537 1 US-08-325-267A-2	Sequence 2, Appli
36	373	6.5	1231 3 US-08-904-263A-4	Sequence 4, Appli
37	371	6.5	761 2 US-08-707-237A-84	Sequence 84, Appli
38	371	6.5	762 1 US-08-642-255-114	Sequence 114, App
39	371	6.5	762 1 US-08-642-255-114	Sequence 26, Appli
40	370.5	6.5	1064 1 US-08-642-255-62	Sequence 62, Appli
41	368	6.5	762 1 US-08-397-633A-26	Sequence 120, App
42	368	6.5	762 1 US-08-397-633A-31	Sequence 31, Appli
43	367.5	6.4	1187 1 US-08-320-559-28	Sequence 28, Appli
44	367.5	6.4	1187 3 US-08-545-860D-28	Sequence 28, Appli
45	367.5	6.4	1187 5 PCT-US94-04496-28	Sequence 28, Appli

ALIGNMENTS

RESULT 1
US-09-103-429A-4
Sequence 4, Application US/09103429A
Patent No. 6187558
GENERAL INFORMATION:
APPLICANT: Granados, Robert R
ATTORNEY/AGENT INFORMATION:
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
TITLE OF INVENTION: C DNA and Related Products and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
STREET: 118 No. 6187558th Tloga
CITY: Ithaca
STATE: NY
COUNTRY: USA
ZIP: 14850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentlin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,429A
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Michaels, Christopher A
REGISTRATION NUMBER: 34,390
REFERENCE/DOCKET NUMBER: BTI-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: (607) 256-2000
TELEFAX: (607) 256-3628
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 805 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoplusia ni
TISSUE TYPE: peritrophic membrane
US-09-103-429A-4
Query Match 9.2%; Score 522; DB 4; Length 805;
Best Local Similarity 27.8%; Pred. No. 76-27;

[illegible]

```

1      CITY: Palo Alto
2      STATE: CA
3      COUNTRY: USA
4      ZIP: 94306-1840
5
6      COMPUTER READABLE FORM:
7      MEDIUM TYPE: Floppy disk
8      COMPUTER: IBM PC compatible
9      OPERATING SYSTEM: PC-DOS/MS-DOS
10     SOFTWARE: Patent In Release #1.0, Version #1.30
11
12     CURRENT APPLICATION DATA:
13     APPLICATION NUMBER: US/08/928,361B
14     FILING DATE: 12-SEP-1997
15
16     CLASSIFICATION:
17     PRIOR APPLICATION DATA:
18     APPLICATION NUMBER: US 60/026,062
19     FILING DATE: 13-SEP-1996
20
21     ATTORNEY/AGENT INFORMATION:
22     NAME: VERNY, Hana
23     REGISTRATION NUMBER: 30,518
24     REFERENCE/DOCKET NUMBER: 480.76-1(HV)
25
26     TELECOMMUNICATION INFORMATION:
27     TELEPHONE: 650-324-1677
28     TELEFAX: 650-324-1678
29
30     INFORMATION FOR SEQ ID NO: 5:
31     SEQUENCE CHARACTERISTICS:
32     LENGTH: 1837 amino acids
33     TYPE: amino acid
34     STRANDEDNESS:
35     TOPOLOGY: linear
36
37     MOLECULE TYPE: protein
38
39     US-08-928-361B-5

```

```

Query Match 508: DB 3: Length 1837:
Best Local Similarity 21.3%; Pred. No. 1.5e-25;
Matches 295; Conservative 100; Mismatches 494; Indels 496; Gaps
48:

QY      33 KVTPTDSTTQGNKSKTSEKPIITAKPINDRPSLPPNSDISKETSLTVNKETTVKEKITT  92
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     307 KHTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT  366
QY      93 TNGKTSYDGEKEKYSKETSKEKTSIETSEKSAKDLPATSKVLAKPTKAEYTTKGPAALTTPKEPT  152
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     367 TTTTITTTTTTKKPPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPTT  426
QY      153 PTTPEKPASTTPKEEPTPTIKSAPTTPKE--APTPTKSAPTKEPAPTTPKEPAPTTP  210
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     427 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT  486
QY      211 KEPAPTTPKEPAATTPKSAPTTPKEPAPTTPKKAPATTPKEPAPTTPKE-----  259
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     487 TTTATTTTTKKPPTTTTTTTTTTTTTTTTTKKTTTTTATTTTTTSETSEVIRKDEMCWLEKNGECE  546
QY      260 -----PTPTT-----PK  266
DB     547 AKGATVGVIGIKDGRIGENGMAPTMIPINDTHVFRERKVKDVGNTLSVRCGKGAGKLEFPD  606
QY      267 EKAPTTPKE-----APTTPKEPA--PTPKPKP-----  291
DB     607 RSLDPIIPVACHNSCSITIVGVSGGCKIHVSPRGSDVSLNSAPIDPCELFNEVYCDTCM  666
QY      292 -----APTTPKEPAPTT--PKPAAPTTPKEPS--PTTPKEPAPTTPKS--  330
DB     667 AKKGATHSGVQTSADSVTTTIAKPTTTTTTGAPQOPATTTTTGSPSKPTTTTTTTKATTTTTT  786
QY      331 ---ADPTTKEPAPTTTKSAPPTTPKEPSPTTTPKEPAPTTPKEPAPTTPKKAPATTPKEPA  386
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     727 LNFITTTTTQKPTTTTTTTTKVCGKRPATTTTTTKPIVTVVTTTKATTTTTTTPPT-----  781
QY      387 PTTPEKPAPTTPKKAP-----TAPKEPA-----  411
DB     782 TTTTTKDEMTTTTTTLPDIDGIEITPTPIEKLMDKTRMTIYDINSGLSDNSDENPIDGSO  841

```

QY 412 -----TTPKETA-PTTP-----KRL 425
Db 842 AGOIAOTSNIPEVQTHKSTGLIPDMVGLPEDPKSGNLVHYTNQTMISGLSVSLAAKNL 901
QY 426 T-----PTPEKLAPTTP-----439
Db 902 TVVDDEYVGLPIDLTGLPDPVSLIPNPETGELFDPISDEIMNGTIAGIVSGISASES 961
QY 440 ----KPAPTTPEE-----LAPTT-----PEEPPTTP--EAPAPTTPKAA 474
Db 962 LLSQKSAIPDPAATMVMYGEFGLNPATGMVIGSLGSPSQTPSPSEIEGIIPEPVAA 1021
QY 475 PNPKEAPPTTPKEAPATTKE-----PA 498
Db 1022 ANADKEKLSIPSVESIPBEKQKIDSISELMYDIESGRLIGOVSKRPIGSIAGDLNPI 1081
QY 499 PTPPKETAPPTPKGTAFTTTLKEAPATTTPKKAPKELAPTTTKEPTSTSKAPATTPKGT 558
Db 1082 MKTPTOTDSVTKRPIDPTT--GLPFPN--PTGHLINPTNNMTMDSFAGAYVAVSNGI 1136
QY 559 APPTT-----PKAPATTTPKEAPATTTPKGTAPTTLKEAPATTTP-----KKPAK-- 601
Db 1137 KTDNVYGLPVDLTGLPKDPSDIPFNSTGELVDPSTGKPINNYAGIYSGKRGLEPIE 1196
QY 602 ----ELAPTTTGP-----TSTT-----SDKRAPTTPEKETAFTTPKEPA-- 636
Db 1197 DEMGNLEPDSSTKLPIDGNNOVLNPETNSVSGSTGSTRKPGIPIVNGGVVDEEAKDQ 1256
QY 637 -----PTTPKRAPATTTPETPTTSEVSTPTTKE--PTTIHKS--PDESTPEL 681
Db 1257 ADKGDGLIAPPTNSINKDOVNTQYSNNTGNINETGKVIPLSGSLNPTSEFTPOQ 1316
QY 682 SAAPTAKALE---NSPKREGVPTTKTPAATKREMTTAAKDKTTERDLRTTPETTTAAPK 738
Db 1317 TDEITGKPDVTVGLPYDPSGTEIIDPAKRLPIGSAVADDELTEVLNITTDVETGLP-I 1375
QY 739 TKEATATTEKTESKITATTTQVSTTTQDTPFKITITLKITTLAKAVTTTKITTE- 797
Db 1376 DLETGLPRDP-----VSLPOLPNCTLYD-----PSNKKPIPGSHS 1411
QY 798 --IMNKPEETAKPKDRATNSKATTPKQAP--TKAP--KKPTSKKPTKTPRVRKKPT-- 849
Db 1412 GFINGTSGEOSHEDKDPSTG-----KPLDPTNGLPEDDESGSLINETDKIQGSHISGF 1465
QY 850 TTPPKMTSTMBELNPTSLIAEAMLOTTTRPNOTPNSKLVENPK-SEDA----- 898
Db 1466 MEVPKPGCEGNGIMTPEOILEAL-----NKLPTSNENVNISPRSSDAVDRPTNTW 1518
QY 899 -----GGAGETPHMLT-----RPHVMPVETPDMOYLPRVNOGIII 936
Db 1519 NKISGOTFQVQDKTIIPGSAASVITHALCTPTQIDPTTGLP--SDPSTGLPPIPGFNVLV 1576
QY 937 NPMLSO-----ETNIC-----NCKPVDGLTTL--RNGTLV-AF 966
Db 1577 DQOTGEQIKGSVPVYSLVYKKNIVTEAAYGLPVDPKTGFPIDPISYLPFAKNGELIDPI 1636
QY 967 RGHYE 971
Db 1637 SGKYF 1641

RESULT 3
5202236-25
; Patent No. 5202236
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSERG,
; SUSAN L.; MCCARDLIS, RUSS; WEI, TERA; PILULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456

; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
; SEQ ID NO: 25
; LENGTH: 744
5202236-25

Query Match 8.9%; Score 506.5; DB 6; Length 744;
Best Local Similarity 29.9%; Pred. No. 6,7e+26;
Matches 239; Conservative 78; Mismatches 308; Indels 175; Gaps 50;

QY 51 PKTTAKPINRPSLPNSDTSKETSILVNNKETVETKETTTTNNQTSIDGKEKTSAKE 110
Db 24 PKMTYPTPKRPSYPP-----YKSKPTLY-----KPKIT----- 53
QY 111 TQSIKTSADLAPNSKVLAKPTPKAETTTKGRPALTTPEEPTTPKE-----PASTTKE 166
Db 54 -----YPTIYK--AKPS-----YPTPKPKKTYPPYKPKLYRPTPKKP 92
QY 167 PPTTIKSAPT--TPKEAPATTTSAPATTTPKEAPATTTPKEAPATTTPKEAP--P 223
Db 93 SYPTPKASPKTYKPKITYPPTYKAKPSYPTPKPKKTYPT--YKRLIYPTPKKASIP 151
QY 224 TTTKASPTTPKEAPATTTPKKAAPT--TPKEAPATT--PKETPTTPKEAPATTKEP-- 275
Db 152 PTYKPKPSYP--PSYKTKTYPTPKPKLYRPTPKRPSYPSYKPKTYPPYKPKLT 209
QY 276 -APTTPEKA--PTAPKKA--PTTPKEBA--PTTPKEBA--PTTKKPS--PTTPKBA 324
Db 210 YPTPKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPS 269
QY 325 -PTTKASAPTTTKEAPATTTSAPTTPEKPSPTTKEBA--PTTPKBA--PTTPK 377
Db 270 YPTPKAKPSYPT--YPTSYKAKPYP--PTYKAKPSYPTPKAKPYKAKPYKAKPYKAK 322
QY 378 A--PTTPKEBA--PTTPKEBA--PTTTPKBA--PTAPKEAPATTTPKEAPTTPKKLTPT 429
Db 323 SYPTPKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPSY 382
QY 430 PEKLAPTPE--KPAATPEELAPTTPEEP--PTTPEBA--PTTPKA--AAPNT--P 478
Db 383 TYKAKPSYPTPKAKPSYPT--PTYKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPS 438
QY 479 KEAPATTTPKEAPATT--PKAPATTTPKEAPATT--PKGAPATTLEAPATTTPKKA 532
Db 439 SYPTPKAKISYPTPKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPSYPTPKAKPS 498
QY 533 ELAPTTTKEPTSTSDKAPATTTPKGTAPTTTPKEAPATTTPKEAPATTTPKKA 591
Db 499 SYPTPKAKPSYPTPKAKPSYPT--ARPSYPT--PTYKAKPS--YPTPKAKPSY 546
QY 552 -PTTPKKAPELAPTTTKEPT--STSDKAPATTTPKEAPATTTPKEAPATTTPKKA--PT 646
Db 547 DPTYKAKPYK--AKPTYKAKPYSTYKAKPSYPTYKAKPSYPT--PTYKAKPSYPT 601
QY 647 ---TPETPTTSEVSTPTTKE-----PTTIHKSPP-----ESTPELSAEP--PKALENS 693
Db 602 YKAKPSYPTPKAKPSYPTYKAKPSYPTYKAKPSYPTYKAKPSYPTYKAKPSYPTYKAK 661
QY 694 PKREGVPTTKPAA-----TKPEMTTAAKDKTTERDLRTTPETTTAAAP-----KMT 739
Db 662 PSYP--PTYKAKPSYPTYKAKPSYPTYKAKPSYPTYKAKPSYPTYKAKPSYPTYKAKPS 719
QY 740 KETATTEKTESKITATTT 759
Db 720 YKAKPYSTYKAKPYST 739

RESULT 4
US-09-103-429A-3

```

Sequence 3, Application US/09103429A
Patent No. 6187558
GENERAL INFORMATION:
APPLICANT: Granados, Robert R
APPLICANT: Wang, Ping
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
TITLE OF INVENTION: cdna and Related Products and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
STREET: 118 No. 6187558th Tloga
CITY: Ithaca
STATE: NY
COUNTRY: USA
ZIP: 14850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,429A
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Michaels, Christopher A
REGISTRATION NUMBER: 34,390
REFERENCE/DOCKET NUMBER: BTI-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: (607) 256-2000
TELEFAX: (607) 256-3628
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 786 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEetical: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoplusia ni
TISSUE TYPE: peritrophic membrane
US-09-103-429A-3

```

```

Query Match      8.7%; Score 498.5; DB 4; Length 786;
Best Local Similarity 29.1%; Pred. No. 2.4e-25;
Matches 200; Conservative 25; Mismatches 207; Indels 255; Gaps 32;
QY 132 PTPKAPETTTGPAITTPKEPTTPPKKAPASTTPKEPTTPPKKAPSTTPKAPSTTPKAPSTTPKAP 191
Db 93 PGPAPAE-TTQAPATTT---QAPTTTQAPTTTT-----QAPTTTQ---ATTTTQAP 135
QY 192 TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 251
Db 136 TTTTQ---APTTTQ---ATTT---QAPTTTQ---APTTTQAPTTTQ---APTTTQ---APTTTQ--- 183
QY 252 PAPTTPKEPTTPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 308
Db 184 -APTTTQAPTTTQAAATTPPATTPPATTPATTPA--TTPATTPPGVAPTPASAPVWPPTIC 241
QY 309 -----PPTTKEPSTTPPK- 321
Db 242 EILPNGCPADFDIHLIPHDKCYKLFQCSNGYTFEORCPGLTFNRYVORCDSPANVEC 301
QY 322 ---EPAPTTT----- 328
Db 302 DGEISAPAPVYEGNEDDIDIDLDNGCPANFEIDWLLPHGNRCDKYYOCVHGNLVERR 361
QY 339 -----KSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 361
Db 362 CGAGTHFSFELQCDHIEIVGCTLPGESEEVVDDEDA-CTGWCYCPTEPIEMERLPGNCP 420

```

```

QY 362 APPT-----PKP-----APTTPKKAPPT----- 381
Db 421 ADFSIDHLLPHESDCOGYLOQCVHGQTARPCGNLHFSPTAOSCESHVTAAGCOVECDSD 480
QY 382 ---KPEAPPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 438
Db 481 NOSTSTAPPTAPPTAPPTAPPTAPPTA--APSTVVPATTPATAPPTT---AIFT-- 533
QY 439 EKPAPTPEELAPPTPEEPPTPEEPAPTTPKAAPT--KPEAPPTTKEPAPTTKE 496
Db 534 --PAPTAPPTAPPTAAPESPTVTVP--PTAAPTAPPTAAPTAPPTVTV--SAPTAPT 587
QY 497 PAPTTPKETAPPTPKGAPTTLKEPAPTTPKKPAKELAPTTTEPSTSDKRAPPTPK 556
Db 588 AAPTAPPTAPPTAAPTAPPTVTVSPTTAPPTAAPT--APNT-----VIVPP 632
QY 557 GPAPTTPKEPAPTTTPKEPAPTTPKGTAPTTTKEPAPTTTPKAPKELAPTTTGPSTTS 616
Db 633 TAPPTT--AAPAPNTVYVPTTAPPTAAPTAPVAA-----APNTTAPVTTTS 677
QY 617 DKPAPTTPKETAPPTTPKEPAPTTPKKP 643
Db 678 ---APATTPEDDIDP--PLPMDPINP 699

```

```

RESULT 5
US-08-700-651-5
Sequence 5, Application US/08700651B
Patent No. 6015882
GENERAL INFORMATION:
APPLICANT: PETERSEN, CAROLYN
APPLICANT: LEECH, JAMES
APPLICANT: NELSON, RICHARD, C.
APPLICANT: GUT, JIRI
TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
FILE REFERENCE: 480.19-4(HV)
CURRENT APPLICATION NUMBER: US/08/700,651B
EARLIER FILING DATE: 1997-08-14
EARLIER APPLICATION NUMBER: 08/415,751
EARLIER FILING DATE: 1995-04-03
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 1721
TYPE: PRT
ORGANISM: Cryptosporidium parvum
US-08-700-651-5

```

```

Query Match      8.6%; Score 489; DB 3; Length 1721;
Best Local Similarity 22.1%; Pred. No. 2.4e-24;
Matches 284; Conservative 82; Mismatches 476; Indels 442; Gaps 41;
QY 5 KKNRTKKKPPKPPVYDEAGSGLDNGDFKYTPDTSTQHNKYSIPKTTAKPINPRS 64
Db 116 RSMETK---TTSPSANYAGVRSN--ETKTTESANT---NFLVDPKI-----N 158
QY 65 LPNDSDTSK-----TSLVNKKETVETKETTNTKQSTNDGKEKTTSAK 109
Db 159 APCSENSFEOGOIFDMGSKVYIPYTKCVGKHHTTTTTTTTTTTTTTTTTT 208
QY 110 ETOSIEKTSKDLAPTSKVLAKPTPKAETTTTCKPALTTTKEPPTTPPKKAPSTTPKEPPT 169
Db 209 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 240
QY 170 TTKSAPTTTKEPAPTTTTSKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 229
Db 241 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 300
QY 230 PTTTKEPAPTTTPKAPATTTTKEPAPTTTPKEPTTPKEPAPTTTKEPAPTTTPKAPATPK 289

```


RESULT 6

[illegible]

```

QY 350 PKE-----PSEPT----- 357
Db 407 TSEIESVIRKPEWCKMLEKNGCEAKGATVYGVIGKDGIEGMAFTMIPNDTHVRFK 466
QY 358 -----TKRP-----APTPKE 368
Db 467 VKDVGNITISVCRKGAGKLEFPDRSLDFTIPVAGHNSCSIIIVGSGDGKIHVSFGSKD 526
QY 369 PAPTTPKRPAPTPKE-----PAPTTPKEPAPTTPKRPAPTPKPA 410
Db 527 VSLIS-----APIOSELEFNEVYCDCTAKYGIHSGYOTSADEVTYTTAKPTTTT--TGA 580
QY 411 PTPKEPAPTTPKRLPTTPPEKLAFTTPPEKPAFTTPPEELAPTPPEEPTTPPEEPAPTTP 470
Db 581 PGPPTTTTSGSPKPTTTTATTTT-----TLNPIITTTTQKPTTTT-----TT- 627
QY 471 KKAAPNTPEKPAFTTPKEPAPTTPKEPAPTTPKETAFTTPKGTATPTTLKEPAPTTPKPA 530
Db 628 --KVFGRPIATTTTTLKRIYTTTTKATTTTTPPTT-----TTTKRDEMPTTTTPL 680
QY 531 PK-----ELAPTTPKEPT-----TSDKPAFTTPKG--TAPTTPKEPAPTTP 570
Db 681 PIGDIEITPIETKMLDKYTMIIYDNGSLLDSDNDPIPSGQAQIADTSLNLPVQTH 740
QY 571 PREPAPTTPKGTAPTTLKE--PAPTTPKRPAPKELAPTTTGPSTTSKPAFTTPKET 627
Db 741 KSTGLPIDMWGLPFDPKSGNLVHPYTNQTMGSLSVYLAANKNLVDDDEYVG--LPIDT 798
QY 628 APPTPEKPAFTTPKRPAPTTPPEPTTPSEVSTPT-----TTKEPTTIKS----- 673
Db 799 LGGYLPIDVSLIPFN--PETGELFDYISDEIMNGTLAGIVSGISASESLISOKSALIDPA 856
QY 674 -----PDESY-----PELSAPTEPKALENSPK 695
Db 857 TNMNVGEFGGLNRPATGVMIPGLPSEQTOFSPEIEDGIIPEVAAANADKFKLSIP- 915
QY 696 EPGVPTTKRPATKPRMTTAKDKTER-----DLRTPPETTTAARPM 738
Db 916 -PSVP-----ESIPR-----KDOKIDSISELMYDIESGRILGOVSKRIPGSIADBLNP 963
QY 739 TKRTATTTTEKTESKITATNT--QVSTSTTQODTTPFKITTLKTTLTAPKVTYTKKTIITTE 797
Db 964 IMKTPOTDSVTSKPIDPTGELFPNPPTGILNPNNNMDSFAAAYKAAVNSNGIKTON 1023
QY 798 IMNKP--EETAKPKD-----RATNSKATTPPKOKPTKAPKPKSTKKPKTMP----- 842
Db 1024 VYGLPVGELGPKKDGSDIPIFNSTGELVDPSTGKPIINNSTAGIVSGKRGLEPIEDENG 1083
QY 843 -----RVKRKKTTPTRKMTS--TMPE----- 862
Db 1084 NLPDPSTNLPIIDGNOLVNEETNSTVSGSTGTTKRPGLIIPVNGGVVDEEAKDOADKG 1143
QY 863 -----INPTSRIAEAMLOTTTRPNQTPNSKLVENPNKSESDAGAEGETPMLLRPHVMP 917
Db 1144 KQGLIIVPTTISINKDPYTNQYSNITGNI--INP--ELGKVIIPGSLPESLNYPSFNP 1197
QY 918 EYTPD-----MDYLPRVNOGIIINP-----MLSETIICNGK----- 950
Db 1198 QQTDEIRGEPVDVYTGIPYPSGTEIIPATKLIPIGSVAGDELIEVLNITTDEVYGLP 1257
QY 951 -----PYOGLTTLRNGTLY 964
Db 1258 IDLETGLPRDVEVSGLPOLPNGTLY 1281

```

```

RESULT 7
US-07-638-431-2
; Sequence 2, Application US/07638431
; Patent No. 5196535
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin

```

```

APPLICANT: Hedstrom, Richard
APPLICANT: Khumsmith, Strin
APPLICANT: Rogers IV, William O.
TITLE OF INVENTION: Protective malaria sporozoite surface protein
TITLE OF INVENTION: Immunogen and gene
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSER: A David Spevack
STREET: NARDC Building 1 T-12 National Naval
STREET: Medical Center
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20814-5044
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/638,431
FILING DATE: 19910110
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, Avrom D.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-4033
TELEFAX: (301) 295-6759
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 826 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-638-431-2

```

```

Query Match 8.6%; Score 488.5; DB 1; Length 826;
Best Local Similarity 26.4%; Pred. No. 1,1e-24;
Matches 153; Conservative 54; Mismatches 223; Indels 149; Gaps 34;

```

```

QY 166 TPTTKASAPTTKEPA--PTTKASAPTTKEPA--PTTKEP-----APTTPKEP--APTTPKE 220
Db 273 TPCKAVHDCPOLIPVIVIPKRIPEKPSNEPEVNPDPNDPNPNPNPNPNPNPNPN 332
QY 221 PAPTTPKSAPTTPKEP--APTTPKEP--APTTPKEP--APTTPKEP--PTTPKEPAPTTPKEPA 276
Db 333 PNNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 392
QY 277 PTPPEKPAFTPAK--KPAFTPEKPAFTTPKPEPAPTTPKEPSPTTPKEPAPTTPKSAFTT 334
Db 393 RNNPKRRNNKPKPKPNKPNKPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 443
QY 335 TKEPAPTTPKSAFTTPKEPSPTTPKEPAPTTPKEPA--PTTPKEP--APTTPKEPAPTTPKE 392
Db 444 PKNPNP-----NEPSNNKPNPN-----EPLNNEPSNPNPNPNPNPNPNPNPNPNPNPN 492
QY 393 PAPTTPKRPAPAPKEP--APTTPKEPAPTTPKRLPTTPPEKLAFTTPPEKPAFTTPPEELAP 451
Db 493 PS-----NPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 523
QY 452 TTPPEEPTTPPEEPAPTTPKAAAPNTPEK--APTTPKEPAPTTPKEPAPTTPKEPAPTTP 510
Db 524 SNPN-----PLNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 564
QY 511 KGTAPTTLKEPAPTTPKRPAPKELAPTTTKEPTSDKPAFTTPKGTAPTTPEKPAFTTP 570
Db 565 -----PSNDEPNPEE--PSNKEP-----SNPEEPIINBELNPKRPSNEEEN 606
QY 571 PKPAPPTTPKGTAPTTLKEPAPTTPKRPAPKELAPTTTGPSTTSKPAFTTPKETAPT 630
Db 607 KKEPI-----NPEESNPKP-----INPDNBNPLIIODEPIEPRNDSNVIDI 649

```


OY	377	PA-PTTPKEBA--PTTPKEBAPTITKKAPAPAKBEAPPTKETAPTTKTLPTTPBK	432
Db	309	PSYPTTKAKSPSYPTTKAKSPSYPTTKAKSPY---PTYKAKSPSYPTTKAKSP--TK	362
OY	433	LAPTPPE--KPAPTPPEELAPTPPEEPT--PTTPEEBA--PTTKAAPAPTKPEBAPTTPK	487
Db	363	AKPYPTSYTKAKSPY---PTYKAKSPYPTTKAKSPYPTTK--AKSPY---PTYKA	413
OY	488	EPA-PTT---PKBPAPTKEETAPTT--PKGAPPTLKBPAPTTPKKAPAKELAPTTPK	540
Db	414	KPYPTSTYKAKSPYPTTKPSTISPYPTTKAKSPSYTKAKSPSYPTTKAKSPYPTTKA	473
OY	541	EPTSTISOKAPPTPKGAPPTPKBAPPTPKBAPPTPKBAPPTKGAPPTLKEBA--PTTPKKP	598
Db	474	KPYPTSYTKAPTYK--AKSPY---PTYKAKS-----YPTTKAKSPYPTTKAKP	521
OY	599	APKELAPPTTKGPTSTSDKAPAPTPKETAAPTTPKEBA--PTTPKKPA--PTT	651
Db	522	TYK--AKPYTKAKSPSYPTTKAKSPY---PTYKAKSPSYPTTKAKSPSYTKAKSPY	576
OY	652	PTTUSVSTPTTKETPTIHKSPDESSTPELSAEPTPKALENSPPKSPVPTTKTPAATKPE	711
Db	577	PTYKAKSPYPTT-----YKAKSPYPTTKAKPTYKA-----KPYPTST--YKAKPS	620
OY	712	MTTAKKOTTERDLATTPETTAAP	736
Db	621	YPTTKAKPYTKAKSPYPTTKAKP	645

```

RESULT 10
US-09-219-849-5
; Sequence 5, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: WOBBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELIE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HAIR EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 960
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
; US-09-219-849-5

```

Query Match	7.9%;	Score 452;	DB 4;	Length 960;
Best Local Similarity	25.6%;	Pred. No. 3.3e-22;		
Matches 173; Conservative	60;	Mismatches 309;	Indels 134;	Gaps 41.

QY 132 PPKAEITTKGPAALTTKEPPTTTPKBEASTTPEKPPPTTIKSAPTTKKEAP---TTTS 189
Db 72 PEPACPEPSGRDG--PPAPGPAG--PGSDDPECPAGC-PAPPCPSRDGPGCAGPA 160
QY 190 APTTEKEPATTTTKEPA-PTTTPKEPATTTTKEPATTTTAKAPTTPKKEPATTP---KKPA 245
Db 127 GPPGSRDDECPGAPCAPGAPCGSDRGPGAGCAPGAPPGSGRDPGGAGAPGPPGCSFDGC 186

```

OY 246 PTTKEPAPTTT---KEPTPTTKEBA - PTTKEAPTTTKEBA----- 284
Db 187 P--PGAPGAPGPGSRDGPBGAPGAPGPGSRDGPBGAGHAGPAGKGAHGAPGAPGAG 244
OY 285 -----PTAPK-KEAPTTT---KEBAPTTTKEBA-PTTKEBSPTTTKEBAPTTSK 330
Db 245 PAGPKGAGPAGPKAGCPGAPGPGSRDGPBGAPGAPGAGPAGPGSRDGP--PGAGP----A 298
OY 331 APPTTKEBAPTTTKSAPTTEKESPTTKEBAPTTEKEBAPTTP---KKAPPTTKEBAP 387
Db 299 GPBGSRDGP-----PGAPGAPGPGSRDGP--PGAPGAPGPGSRDGP--PGAPG 348
OY 388 TTP---KEBAPTITKKP-----PTAPKKEBAPTTEKETAPTTPKLPTTEKLAPT 437
Db 349 AGPPGSRDGPBGAPGAPGPGSRDGPBGAPGAPGAGPAGPGSRDGP-----PGAGCAPGPG 404
OY 438 PEKRAPT--TPELAPTTPEEPTPTTPEEAPTTTAKAA--PNTKEBAPTTP---KEBAPT 492
Db 405 SMDPPGAPGAPGAGPAGSRDGP--PGAPGAPGPGSRDGPBGAPGAPGPGSRDGP 462
OY 493 TPEKA-----PTTRET-APTTPKGT-----APTTLKAPATTTKK--APRELA 535
Db 463 GAHGAPGAPGAGPAGPAGKAGAPGKGAHGAPGAPGAGPAGPAGPGSRDGPBGAPGAG 522
OY 536 PTTTKEPSTSDKAPTTTKGIA-PTTKEBAPTTP---KEBAPTTKGA-PTTKEP 590
Db 523 PGGSRDGPBGAP--PGAPGPGSRDGPBGAPGAPGPGSRDGPBGAPGAPGAGPAGPGSRD 580
OY 591 APPTTKKRP-----KEIAPTTTGPST--TSDKAPTTTKEBAPTTEKEBAPTTPK 642
Db 581 GP--PGAPGAPGPGSRDGPBGAPGAGPAGPGSRDGPBGAPGAGPAGPGSRDGP--PGA 636
OY 643 PAPTP---ETPPPTSEVSTPTTKEPTTIHKSPESTELSAEPTKALEN-----S 693
Db 637 PGAGPAGPGSRDGPBGAPGAPGPGSRDGP--PGAPGAPGPGSRDGPBGAPGAPGPG 695
OY 694 PKEPGPTTKTPAATK 709
Db 696 SRDGPBGAGPAGPK 711

```

RESULT 11
US-08-479-537A-5
Sequence 5, Application US/08479537A
Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMSON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HARDEVINT, Maria
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 2213-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 51A
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teekin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-08-479-537A-5

Query Match 7.4%; Score 424.5; DB 2; Length 1867;
Best Local Similarity 23.7%; Pred. No. 4.5e-20;
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;
75 TSLTVKETTVEKETTNTKOTSDGKETSAAKETOSIEKTSKD-LAPTSKVLAKPT 133
16 TVLTIV-----VTGSHASSTPGGKETSATQRSSVPSSTKNAVSMSSVLSHS 65
QY 134 P-KAETTTKGP--ALTTPKP-----TPTTPKEPASTTPKEPTPTIKSAP 176
DB 66 PGGSGSTTGGDVTFLAPATEPASGAATWGDVTSVPTVRPALGSTTPRAHDVT---SAP 122
QY 177 TTPKEPAPTT-----KSAPTTPKEP-----APTTPKEPAPTTTPKEPAPT 217
DB 123 --DNKPAPEGSTAPXAHGVTSA PDKRPXGSGTAPXAHGVTSA PDKRPXGSGTAPXAHGVT 180
QY 218 TKE--PAPTTT-----KSAPTTPKEPAPTTTPK-----KPAPTTPKEPAPTTTPKPTPTT 264

```

```

DB 181 APDXRPXGSGTAPXAHGVTSA PDKRPXGSGTAPXAHGVTSA PDKRPXGSGTAPXAHGVT 240
QY 265 PKE--PAPTTKEP-----APTTPKEPAPTAK-----KPAPTTPKEPAPTTPKEPAPT 311
DB 241 APDXRPXGSGTAPXAHGVTSA PDKRPXGSGTAPXAHGVTSA PDKRPXGSGTAPXAHGVT 300
QY 312 TKEPSTTPKEPAPT--TKSAPTTPKEPAPT-----TKSAPTTPKEPAPTTPK-----E 360
DB 301 APDXRP--XPGSTAPXAHGVTSA PDKRPXGSGTAPXAHGVTSA PDKRPXGSGTAPXAHGVT 359
QY 361 PAPTTTPKEPAPTTPK-----KPAPTTPKEPAPTTPK-----EPAPTTPKKAAPTAPKEPAPT 412
DB 360 SAPDXRPXGSGTAPXAHGVTSA PDKRPXGSGTAPXAHGVTSA PDKRPXGSGTAPXAHGVT 419
QY 413 TPKETAP---TTPKRLTPTTPKEKLAPTTPKEKAPPTTPPELAPTTPEEPTTPTEPAPT 468
DB 420 SAPDXRPXGSGTAPXAHGVTSA---APDXRPXGSGTAPXAHGVTSA PDKRPX---PGST 471
QY 469 TPKA---AAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 523
DB 472 APXAHGVTSA PDKRPXGSGTAPXAHGVTSA PDKRPX---XPGSTAPXAHGVTSA PDKRPXGSG 530
QY 524 TTP-----KKPAPELAPTTTKEPTSTTSOKRAP--TTPKGTAPTTTPKEPAPTTPTE 571
DB 531 TAPXAHGVTSA PDKRPXGSGTAP--XAHGVTSA PDKRPXGSGTAPXAHGVTSA---APDXR 585
QY 572 KEPAPTTPKG---TAPTTLKEPAPTTP-----KKPAPELAPTTTGTSTTSD 617
DB 586 PXRGPSTAPXAHGVTSA PDKRPXGSGTAPXAHGVTSA PDKRPXGSGTAP--XAHGVTSA PDX 644
QY 618 KPAP--TTPK-----EPAPTTPKEPAPTTP-----KKPAPTTPETTPPTSEVSTP 661
DB 645 KRPXGSTAPXAHGVTSA PDKRPXGSGTAPXAHGVTSA PDKRPXGSGTAP--TAPXAHGVTSA 702
QY 662 TTPKEPTT-----IHKSPDESTPELSAEPTRKALENSPKRPGVPTTKTPA----- 707
DB 703 DXRPXGSGTAPXAHGVTSA PDKRPXGSGTAPXAHGVTSA PDKRPXGSGTAPXAHGVTSA 762
QY 708 -TKPEKTTTAKK---TTERDLRTTPETTTAPKMTKETATTTETKTSKITATTTQVTS 763
DB 763 DXRPXGSGTAPXAHGVTSA PDKRPXGSGTAP--APXAHGVTSA PDKRPXGSGTAPXAHGVT 820
QY 764 TTTQDTPPEKITTLKTTTLAPRVTTTKKTTITTTTELMMKPEELAKKDKATTSKATTPKQ 823
DB 821 APDXRPXG---STAPXAHGVT-----SAPDXRPXGSGTAPXAHGVTSA PDKRPX-- 868
QY 824 KPTKAPKRPSTSKRPKTPRVAKRPKTPTPPKMTSTMBELNPTSTIAEAMLOTTTRPQ 883
DB 869 -GSTAPXAHGVTSA PDKRP--XPGSTAPXAHGVTSA--PDKRPX-----PGST 911
QY 884 -PNSKLVEVNKSEEDGAGAEGETPMMLRPVHVEEVTPTMDYLPRVNOGIIINPMLSD 942
DB 912 APXAHGVTSA PDKRPXGSGTAPXAH-----GVTSA PDKRPXGSGTAP--PXAHG 957
QY 943 ETNINCKRPVDTLT--LRNGTLVAFRGHYFMMLSPFSPSPARRTTEWGIPTSPID 997
DB 998 VTSAPDXRPXGSGTAPXAHGVTSA-----PDKRPXGSGTAPXAHGVTSA PD 1003

RESULT 12
US-09-083-116-5
Sequence 5, Application US/09083116
Patent No. 6203795
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENVY, Marie-Paule
APPLICANT: LAIHE, Richard
APPLICANT: HAREUVENI, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
NUMBER OF SEQUENCES: 5
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHEIS, L.L.P.

```

STREET: P.O. Box 1404
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: United States
 ZIP: 22313-1404
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/083,116
 FILING DATE:
 CLASSIFICATION:
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: 08/479,537
 FILING DATE:
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/FR91/00835
 FILING DATE: 23-OCT-1991
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/039,320
 FILING DATE: 04-APR-1993
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/403,576
 FILING DATE: 14-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Teeklin, Robin L.
 REGISTRATION NUMBER: 35,030
 REFERENCE/DOCKET NUMBER: 017753-025
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1867 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 128..1727
 OTHER INFORMATION: /note= "The amino acids spanning 128 to 1727 constitute a repeated region wherein the repeat is 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."
 OTHER INFORMATION:
 NAME/KEY: Peptide
 LOCATION: 134
 OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
 OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 144
 OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACG
 OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 147
 OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
 OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
 NAME/KEY: Peptide
 LOCATION: 1..21
 OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
 OTHER INFORMATION: 21 amino acid precursor sequence."
 US-09-083-116-5

Query Match 7.4% Score 424.5; DB 4; Length 1867;
 Best Local Similarity 23.7%; Pred. No. 4.5e-20;
 Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY	75	TSLVNKEETVEETKETTNTKQTSIDGKEKTSKETSQSIETKTSKD-LAFTSVLAKPT	133
DB	16	TVLVY-----VTSGHASSTPGGEKETSATQSSSVSSTEKNVSMSTSVLSHS	65
QY	134	P-KAETTKGP--ALTTPKEP-----TTPTPKEPSTTPKEPTTTPKAP	176
DB	66	PGSGSTTGQGDVTLAPAPETPAGSGSAATWGQDVTSVPTRALDSTTPPAHDVY---SAP	122
QY	177	TTPKRPAPPTT-----KSAPPTTPKEP-----APTTPKEPAPPTTPKEPAPT	217
DB	123	--DNKRPAGOSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVT	180
QY	218	TKE--PAPTT-----KSAPPTTPKEPAPPTPK-----KRAPTPPKEPAPPTPKEPPTT	264
DB	181	APDXRPXPGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVT	240
QY	265	PKE--PAPTPKEP-----APTTPKEPAPTPK-----KRAPTPPKEPAPPTTPKEPAPT	311
DB	241	APDXRPXPGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVT	300
QY	312	TKEPSEPTTPKEPAPPT--TKSAPPTTPKEPAPT-----TKSAPPTPKESPPTTK-----E	360
DB	301	APDXRP--XPGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVT	359
QY	361	PAPTPPKEPAPPTPK-----KRAPTPPKEPAPPTPK-----EPAPPTTKKRAPTPAPKEPAPT	412
DB	360	SAPDXRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVT	419
QY	413	TPKETAP-----TTPKKIAPPTPEKIAPTTPKEPAPPTPEKIAPTTPPEEPTTPPEEAPT	468
DB	420	SAPDXRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVT	471
QY	469	TPKA-----AAPTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPT	523
DB	472	APXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAP	530
QY	524	TTP-----KKRAPKELAPTTTPKEPAPPTTPKEPAPPTTPKEPAPPTTPKEPAPT	571
DB	531	TAPXAHGVTSAADKRPXGSTAP--XAHGVTSAADKRPXGSTAPXAHGVT-----APDXR	585
QY	572	KEPAPPTPKG---TAPPTLKEPAPPTP-----KKRAPKELAPPTTPGPTSTSD	617
DB	586	RPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAP--XAHGVTSAADK	644
QY	618	KRAP--TTPK-----ETAPTPKEPAPPTP-----KKRAPTPPEPPTTSEVSTP	661
DB	645	RPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAP--TAPXAHGVTSA	702
QY	662	TTPKEPTT-----IHKSPDESTPELSAEPTRKALENSPKEGVTTPTPAA-----	707
DB	703	DXRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSA	762
QY	708	--TRPEMTTKKDK---TTEEDLTTPETTTAAKMKKEATTEKTESKITATTTQVTS	763
DB	763	DXRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVT	820
QY	764	TTQDTTPFKITLTKTTTLAPKVTYTKKITTTEIMNKKPEETAKPKDRATNSKATTPKQ	823
DB	821	APDXRPXG-----STAPXAHGVT-----SADDXRPXGSTAPXAHGVTSAADKRPX	868
QY	824	KTPAPKPKPTSTKKPKTPPVKRPKTTTPTRKMTSTMPLELNPSTSLAEAMLQTTTPRNO	883
DB	869	--GSTAAPXAHGVTSAADKRP--XPGSTAPXAHGVTSA--PDXRPX-----PGST	911
QY	884	--PNSKLVEVNPKSDAGAGETPRHMLLRPHVMEPEVTTPMDLPRVNOGIIINPLSD	942
DB	912	APXAHGVTSAADKRPXGSTAPXAH-----GVTSAADKRPXGSTAP--PXA	957
QY	943	ETNICNGKPVDLGTT--LRNGTLVAFRGHTFMMLSPPSPSPARRITTEVWGISPID	997

Db 958 VTSAPDXRXPXGSTAPXAHGVTSA-----PDXRXPXGSTAPXAHGVTAPD 1003

RESULT 13

US-08-479-537A-2

Sequence 2, Application US/08479537A

Patent No. 5861381

GENERAL INFORMATION:

APPLICANT: CHAMBER, Pierre

APPLICANT: KIENY, Marie-Paule

APPLICANT: LAFRE, Richard

APPLICANT: HAREVENT, MIRA

TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE

TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/479,537A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 90/13101

FILING DATE: 23-OCT-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/FR91/00835

FILING DATE: 23-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/039,320

FILING DATE: 04-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/403,576

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Rodin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 017753-025

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2035 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Peptide

LOCATION: 128..1899

OTHER INFORMATION: /note="The amino acids spanning

OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat

OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such

OTHER INFORMATION: repeats varies from 1 to 40."

FEATURE:

NAME/KEY: Peptide

LOCATION: 134

OTHER INFORMATION: /note="Amino acid 134 is X1 - Xaa

OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,

OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 144

OTHER INFORMATION: /note="Amino acid 144 is Y - Xaa

OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,

OTHER INFORMATION: or ACG; and Asn = AAT or AAC."

FEATURE:

NAME/KEY: Peptide

LOCATION: 147

OTHER INFORMATION: /note="Amino acid 147 is X2 - Xaa

OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,

OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 1..21

OTHER INFORMATION: /note="Amino acids 1 to 21 are a

OTHER INFORMATION: 21 amino acid precursor sequence."

US-08-479-537A-2

Query Match 7.4%; Score 424.5; DB 2; Length 2035;

Best Local Similarity 23.7%; Pred No. Se-20;

Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

Db 75 TSLVNETTETVETKETTNNKQSTDGKEKTSARETSIETKSAD-LAPTSVLAKPT 133

Db 16 TVLTV-----VTGSGHASTTGGCKERSATGRRSSVPSSTKNAVSMTSSVLSHS 65

Db 134 P-KAETTKGP--ALTTPK-----TPTPKBPASTTKEPTPTTIKAP 176

Db 66 PGGSSSTTGGODVTLAPATEPASGAATWGODVTSVPTRALPSTTPPAHDVT---SAP 122

Db 177 TTPKEPAPTT-----KSAPTTKPEP-----APTTTKEPAPTTKPEPAPTT 217

Db 123 --DNKPAFGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVT 180

Db 218 TKE--PAPTT-----KSAPTTKPEPAPTTK-----KPAPTTKEPAPTTKPEPTT 264

Db 181 APDXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVT 240

Db 265 PKE--PAPTTKPEP-----APTTPKEPAPTPK-----KPAPTTKEPAPTTKPEPAPTT 311

Db 241 APDXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVT 300

Db 312 TKEPPTTKEPAPTT--TKSAPTTTKEPAPTT-----TKSAPTTKPEPSTTTK-----E 360

Db 301 APDXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVT 359

Db 361 PAPTTKPEPAPTTK-----KPAPTTKEPAPTTK-----EPAPTTTTPKAPTAKEPAPT 412

Db 360 SAPDXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVT 419

Db 413 TPKETAP--TTPKTLPTTPKELAPTTPEKRAPTTPEELAPTTPEEPPTTPEEPAPT 468

Db 420 SAPDXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVT 471

Db 469 TPKA-----AAPTTKEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPGT--APTTLKBPAP 523

Db 472 APXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXG 530

Db 524 TTP-----KKPAKELAPTTTKEPTSTSDKAP--TTPKGTAPTTKPEPAPTT 571

Db 531 TAPXAHGVTSAPODXRXPXGSTAP--XAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXP 585

Db 572 KEPAPTTPKG-----TAPTLTKEPAPTT-----KKPAKELAPTTTKEPTSTSD 617

Db 586 PXPSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAP--XAHGVTSAPODX 644

Db 618 KPAP--TTPK-----ETAPTTKPEPAPTT-----KKPAPTTPEPTTSEVSTP 661

Db 645 RXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXG--TAPXAHGVTSAPO 702

Db 662 TTPKEPT-----HKSPDESTPELSAEPTRKALENSKEEGVPTTKTPAA-----707

Db 703 DXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPODXRXPXGSTAPXAHGVTSAPO 762

```

OY 708 -TKPEMTTAKD---TTERDLRTTPEYTTAAPKMTKETATTEKTESKITATTQVTS 763
DB 763 DAPRXPSTAPXAHGVTAPDXRXPXGST--APXAHGVTAPDXRXPXGSTAPXAHGVT 820
OY 764 TTTQDTTPPKITLTKTTLAPVYTTTKITTTTEIMNKKEETAKRKDRATNSKATTPKQ 823
DB 821 APDXRXPXG-----STAPXAHGVT-----SAPDXRXPXGSTAPXAHGVTAPDXRXPX 868
OY 824 KPTKAPKPTSTTKPKTMRVRRKPTTPPRKMTSTMPELNPTSLAEAMLQTTTRPNOT 883
DB 869 -GSTAPXAHGVTAPDXRXPXGSTAPXAHGVTSA-PDXRXPXGSTAPXAHGVTSA-----PGST 911
OY 884 -PNSKLVENPXSSEDAGAEGETPHMLLRPHVPMPEVTPDMQTLPRVPMQIINPMUSD 942
DB 912 APXAHGVTAPDXRXPXGSTAPXAHGVTSA-----GVTAPDXRXPXGSTAPXAHGVTSA-----PXAHG 957
OY 943 ETNICKGRKPYDGLTT-LRNGTLVAFRGHYFWMLSPPSPSPARLIEVWGIPSPID 997
DB 958 VTSAPDXRXPXGSTAPXAHGVTSA-----PDXRXPXGSTAPXAHGVTSA 1003

RESULT 14
US-09-083-116-2
Sequence 2, Application US/09083116
Patent No. 6203795
GENERAL INFORMATION:
APPLICANT: CHAMRON, Pierre
APPLICANT: KIENEY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUVENT, Maira
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TITILE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
City: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,116
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Testin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid

```

```

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repea
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X = Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CC
OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1.21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-2

Query Match 7.4%; Score 424.5; DB 4; Length 2035;
Best Local Similarity 23.7%; Pred. No. 5e-20; Indels 241; Gaps 53;
Matches 255; Conservative 69; Mismatches 511;
OY 75 TSLTVNKETTVERKETTNTKSTGDKETTSARETOSIEKTSKD-LAPTSKIYAKPT 133
DB 16 TVLTIV-----VTGSGHASSSTPGGKERKTSATQGRSSVPSSTENAVSMSTSVLSHS 65
OY 134 P-KAEITTKGP--ALTTPKEP-----TPTPKKEPASTTPKEPTTITKSA 176
DB 66 PGSGSSTQGOQDVTLAPATEPASGSAATWGDVTSVPTRPALGSTTPPAHDVT--SAP 122
OY 177 TTPKEPAPTT-----KSAPTTKEP-----APTTPKEPAPTTKEPAPTT 217
DB 123 --DNKRAPGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 180
OY 218 TKE--PAPTT-----KSAPTTKEPAPTTPK-----KPAPTTKEPAPTTKEPAPTT 264
DB 181 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 240
OY 265 PKE--PAPTTKEP-----APTTPKEPAPTTAK-----KPAPTTKEPAPTTKEPAPTT 311
DB 241 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 300
OY 312 TKEPSPTTPKEPAPTT--TKSAPTTTKEPAPTT-----TKSAPTTTKEPSPTTTK---E 360
DB 301 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 359
OY 361 PAPTTKEPAPTTPK-----KPAPTTKEPAPTTPK-----EPAPTTTKKAPAPTAPEAPPT 412
DB 360 SAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 419
OY 413 TPKETAP-----TTPKILTPPTPKIAPTTPKEKAPPTTPELATTTPEEPTTPPEAPPT 468
DB 420 SAPDXRXPXGSTAPXAHGVTSA-----APDXRXPXGSTAPXAHGVTAPDXRXPXGST 471
OY 469 TPKA-----AAPNTPKPEPAPTTTPKEPAPTTTPKEPAPTTTPKETATTPTKGT-APTTLKEPAP 523

```



```

Db 472 APXAHGVTSAADKRPXGSTAPXAHGVTSAADKRP-XPGSTAPXAHGVTSAADKRPXG 530
QY 524 TTP-----KKPAKELAPTTTKEPTSTSDKAP--TTPKGAAPTTKEPAATTP 571
Db 531 TAPXAHGVTSAADKRPXGSTAP-XAHGVTSAADKRPXGSTAPXAHGVTSA--APDXR 585
QY 572 KEPAATTPKGC---TAPTTKEPAATTP-----KKPAKELAPTTTKEPTSTSD 617
Db 586 PXPGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAP-XAHGVTSAADK 644
QY 618 KPAK--TTPK-----ETAPTTKEPAATTP-----KKPAATTPETTPPTTSEVSTP 661
Db 645 RXPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGST--TAPXAHGVTSA 702
QY 662 TTPKEPT-----IHKSPDESTPELSAPETPKALENSKEGCVTTTKTPAA----- 707
Db 703 DXAPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSA 762
QY 708 -TTPKMTTAKDK--TTERDLRTTPEYTTAARPKMTKETATTEKTESKITATTTQVTS 763
Db 763 DXAPXGSTAPXAHGVTSAADKRPXGST--APXAHGVTSAADKRPXGSTAPXAHGVT 820
QY 764 TTPQDTPPEKITLKTTLAPKVTTTTKKTTTTEIMNKPEETAKPKDRATNSKATTPKQ 823
Db 821 APXKRPXG-----STAPXAHGVT-----SAPDXRPXGSTAPXAHGVTSAADKRPX 868
QY 824 KPTAKRKPTSTKKPKTMRVVRKPTTPRKMTSTMPLEINPTSLAEMLQTTTPNQT 883
Db 869 -GSTAPXAHGVTSAADKRPXG--XPGSTAPXAHGVTSA--PDXRPX-----PGST 911
QY 884 -PNSKLYENPNSSEDAGLEGETPHMLLRPHVFMPEVTPDMVLPRLVNPQGIILINPMLSD 942
Db 912 APXAHGVTSAADKRPXGSTAPXAH-----GVTSAPXKRPXGSTA--PXAHG 957
QY 943 ETNICKGRKPVDTLT--LRNGTLVAFRGHFMMLSPSPSPARITLEVIGISPID 997
Db 958 VTSAPDXRPXGSTAPXAHGVTSA-----PDXRPXGSTAPXAHGVTSAAD 1003

```

RESULT 15

US-08-276-967-2

Sequence 2, Application US/08276967

Patent No. 5851817

GENERAL INFORMATION:

APPLICANT: Hardy, Daniel M.

APPLICANT: Garbers, David L.

TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of

TITLE OF INVENTION: Sperm

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P. O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/276,967

FILING DATE: Submitted Herewith

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Kitchell, Barbara S.

REGISTRATION NUMBER: 33,928

REFERENCE/DOCKET NUMBER: UTSD:418\KIT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 713-787-1400

TELEFAX: 713-789-2679

```

; TELX: 79-0924
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2476 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-276-967-2

```

Query Match 7.4%; Score 419.5; DB 2; Length 2476;
Best Local Similarity 32.5%; Pred. No. 1,3e-19;
Matches 163; Conservative 43; Mismatches 177; Indels 119; Gaps 31;

```

QY 406 KEPAATTPKEPAATTPKKTTPTEPKLAPTTPEKAPATTPPELAATTPTEEPPTTPEEP 465
Db 313 PSETSVSTKPAAPTE---KPLVPEIYITPEKPMVMKEKPIVHT--EKPTVPT-EKP 365
QY 466 APPTPAAAPNTPKEPA---PTTKREPAATTPKEPAATTPKEPAATTPKGAATTLKEP 521
Db 366 TTPTEKSTVPT--KKPTVKEPTLPPE-GPTVPAE-RPTTPEGPAVPEKG--PTVLE- 418
QY 522 APPTPKPAKELAPTTTKEPTSTSDKAPATTPPGTAPT--TPKEPAATTPKEPAATTPK 580
Db 419 -----WPSHTEKSTVHTKELIPTGKSTIPTKEPMVPTKRT-----TTP- 458
QY 581 GAAPTTLKEPAATTPKPAKELAPT--TTKPTSTT---SDKPAATTPKEPAATTPKPEPA 636
Db 459 -TERTTIPAEKPTV---PIEKPMVPTERTTIPTEKPTTIPTEKPIVPTKELVPT--EKPI 512
QY 637 PTPPKPAATTPTEPP-----TSEVSTPT-----TKKEPTTIKSDSESTPELSAEP 686
Db 513 VTEKPIVPTKELVPTTEKLVLEERTTTPERTTIPTEKPIVPEKPSVPT-EKPIVPT 571
QY 687 PKALENSKEGCVPTTKTPAATKPEMTTAKDTERDLRTTPEYTTAARPKMTKETATTT 746
Db 572 -----EEPTIPEKELVPT--ERTTTPKRTTTPTRITTPTI-----RTTPT 613
QY 747 EKTESKITATTTQVTSSTTODTTPFKITTLKTTTLAPKVVTKKTTTTEIMNKPEETA 806
Db 614 ERTT-----PTTITTTTPTERTT--IPTKKT-----VPEKTIPT-----ERTI 652
QY 807 KPKDRATNSKATTPKPKOKPTAPKPKPTSTKKPKTMRVVRKKTTPTPKMTSTMPLEINPT 866
Db 653 AP-----TTPQP-SPTLVPTQPAVVMPEST-----SATVTPPTTIAACP---PN 693
QY 867 SRIAEAMLQTTTRPNQTPNSKL 888
Db 694 AHFERCACPVSCQ--SPTPNCEL 714

```

Search completed: April 26, 2002, 16:17:20
Job time: 231 sec

